CORRELATION OF EEG RESPONSES REFLECTS STRUCTURAL SIMILARITY OF CHORUSES IN POPULAR MUSIC

Neha Rajagopalan
Stanford University
nehara j@stanford.edu

Blair Kaneshiro
Stanford University
blairbo@stanford.edu

ABSTRACT

Music structure analysis is a core topic in Music Information Retrieval and could be advanced through the inclusion of new data modalities. In this study we consider neural correlates of music structure processing using popular music—specifically choruses of Bollywood songs—and the NMED-H electroencephalographic (EEG) dataset. Motivated by recent findings that listeners' EEG responses correlate when hearing a shared music stimulus, we investigate whether responses correlate not only within single choruses but across pairs of chorus instances as well. We find statistically significant correlations within and across several chorus instances, suggesting that brain responses synchronize across structurally matched music segments even if they are not contextually or acoustically identical. Correlations were only occasionally higher within than across choruses. Our findings advance the state of the art of naturalistic music neuroscience, while also highlighting a novel approach for further studies of music structure analysis and audio understanding more broadly.

1. INTRODUCTION

Music structure analysis (MSA)—the task of dividing and labelling songs into perceptually salient segments [1]—is a core topic of Music Information Retrieval (MIR) and has been approached through a variety of data types including audio representations, lyrics, and perceptual annotations. For example, choruses of popular songs are often easily recognizable by music listeners, and can be detected from audio due to both their placement throughout a song and their intrinsic features [2]. While much progress has been made in this area, there may be new approaches and data modalities that could advance it even further.

MIR studies have come to involve brain data, particularly electroencephalography (EEG) [3]. EEG has been used to predict stimulus labels, decode musical attributes such as beat and tempo, and even reconstruct music. EEG inter-subject correlation (ISC), which captures neural synchronization of audience members experiencing a complex, real-world stimulus [4], has also advanced music neuroscience research. We leverage and extend this approach to investigate MSA.

We focus on responses to four Bollywood songs written in the popular form—specifically their choruses, due to their salience and tendency to repeat with a high degree of similarity. Importantly, while past EEG-ISC studies have considered responses among listeners experiencing the same stimulus (e.g., one chorus instance), we ask for the first time whether neural responses also synchronize *across* instances of structurally similar content (i.e., pairs of choruses). Moreover, by using a dataset containing two response trials from each participant, we can investigate correlations both across and within participants. In sum, we address the following research questions:

RQ 1 *Does music structure similarity translate to measurable similarity among responses?* In other words, do brain responses synchronize across structurally matched musical segments, even when those segments are contextually unique (in their placement within the song) and also often acoustically unique from one another? Here we expect structural similarity to produce statistically significant EEG correlations both within and across a song's choruses.

RQ 2 *Even if responses are similar across chorus instances, are individual choruses still uniquely experienced?* This question extends RQ1 to investigate whether EEG responses are more correlated within, versus across, chorus instances. We predict that within-chorus EEG correlations will be higher than across-chorus correlations.

RQ 3 *Are a listener's neural responses more similar to themselves than to responses from other listeners?* Understanding whether reliable measures of music structure similarity can be obtained from single listeners can motivate the design of future studies. We expect EEG correlation with one's own data will be higher than correlation with the data of other listeners, due to individual differences in perception and EEG characteristics.

We report small but often significant correlations that align with previous published research. Moreover, within-chorus correlations do not systematically outperform across-chorus correlations. While preliminary, our findings suggest that this novel application of EEG correlation may capture structural similarity during music listening, which may motivate future MSA studies.

2. RELATED WORK

2.1 MSA and Chorus Analysis

MSA involves recognizing and labelling non-overlapping musical segments based on musical similarity [1]. Over the years, MSA has come to involve specific features, similarity representations, and algorithms [5]. One sub-topic of MSA is chorus identification; here, choruses have often been identified based on repetition and contextual cues using measures of similarity [6] and Markov models [7], as well as chroma features and image processing filters [8]. Some systems have also used segment length and positioning to identify choruses [6, 8]. Independently of context, Van Balen et al. looked at intrinsic content features that might distinguish choruses [2]. Their "Chorusness" variable, a probability measure of how likely a segment may be labelled as a chorus by an independent annotator, highlights audio features (e.g., higher loudness and roughness) that qualify the particular salience of choruses.

MSA remains a challenging task due, for example, to ambiguities around defining similarity as well as subjectivity and interpretation of annotations [1,9]. In their 2020 overview article, Nieto et al. called for "richer human labels in upcoming MSA datasets" [1]; we propose that brain data may fit this call.

2.2 MIR and EEG

The growing use of decoding and signal-based approaches and complex, naturalistic (real-world) stimuli in neuroscience has increased that field's relevance to the more applied field of MIR. Kaneshiro & Dmochowski have suggested that MIR and neuroscience researchers might augment their gains through collaboration, highlighting EEG as a particularly relevant response type for MIR due to its high temporal resolution, non-invasiveness, whole-brain coverage, and relative portability and low cost [3].

EEG studies addressing MIR topics include using classification to predict which stimulus elicited an EEG response [10, 11] or which stream a listener attended to in a polyphonic stimulus [12–15]. Other tasks include EEG-based tempo detection/classification [16–18], onset detection [19], and music reconstruction [20]. EEG has been mapped to time-varying music or audio features using Canonical Correlation Analysis (CCA) [21] or deep-CCA [22]; by correlating EEG with semantic music vectors [23]; or using MEG—the magnetic analogue of EEG—and temporal response functions to decode surprisal [24]. In recent studies, Ofner and Stober examined EEG responses at automated segmentation boundaries [25], and Sangnark et al. performed music preference classification on EEG responses to choruses with and without lyrics [26]. However, we know of no study to date that has assessed *similarity* among EEG responses to repeated structural segments.

2.3 Neural Correlation

A particularly relevant approach for the current study involves the correlation of neural responses to a shared stimulus, often termed inter-subject correlation (ISC). Hasson et al.'s 2004 seminal functional magnetic resonance imaging study showed that real-world stimuli (e.g., movie excerpts) can synchronize neural responses across audience members, and that the timing and location of synchronized activity identifies stimulus-evoked brain activity [27]. This data-driven approach, reducing the need for controlled stimuli and a priori event markers, facilitated the use of complex stimuli in neuroscience. In 2012, Dmochowski et al. introduced an EEG implementation which first optimizes the data for ISC [4]. Often referred to as "Correlated Components Analysis (CorrCA)" [4] or "Reliable Components Analysis (RCA)" [28], this optimization applies a relative eigenvalue decomposition to compute multiple spatial filters in which across-trials variance relative to within-trials variance (i.e., ISC) is maximized.

Recent studies involving music have shown that EEG-ISC is modulated by listener expertise [29], musical tempo [30], temporal stimulus manipulations [30, 31], and salient musical events [31]. Auditory studies have reported small but significant group-mean ISC ($0.01 < r < 0.02$) in RC1, the maximally reliable spatial component. Repetition, explored through repeated listens of full excerpts, sometimes but not always results in lower ISC on repeated listens [29, 30]. However, the topic of repeating structural elements *within* a song has not yet been addressed.

2.4 Music-EEG Datasets

The acquisition and preparation of EEG data for analysis requires specialized expertise and sizeable investments in recording apparatus [3]. A key factor supporting MIR-EEG research is the growing number of open EEG datasets released with the intent for re-use by other researchers. Datasets vary in stimuli, stimulus manipulations, participant samples, listening tasks, additional response types, and EEG platforms used. Shorter stimuli are used in the MIIR dataset, comprising perceived and imagined responses to 12 excerpts 6.9–16.0 seconds in length [32] and the MAD-EEG dataset involving 78 solo, duet, or trio stimuli, each around six seconds long [14]. Datasets involving slightly longer excerpts include the DEAP dataset, with 40 one-minute excerpts from music videos [33]; MUSIN-G, with 12 excerpts, 100–132 seconds in length, from various genres [34]; and NMED-M, containing five-minute excerpts of various versions of a minimalist work [31]. Finally, a few datasets use complete musical works as stimuli: NMED-H includes four Bollywood songs [35], NMED-T uses 10 EDM-style songs [36], and NMED-E includes a cello concerto movement [37].

3. METHODS

3.1 EEG Dataset and Stimuli

Among the available datasets, we chose to work with NMED-H (Naturalistic Music EEG Dataset—Hindi) [35] as it used full-length pop (Bollywood) songs with repeating choruses as stimuli. Specifically, we work with the four "Intact" songs of the dataset: "Ainvayi Ainvayi", "Daaru

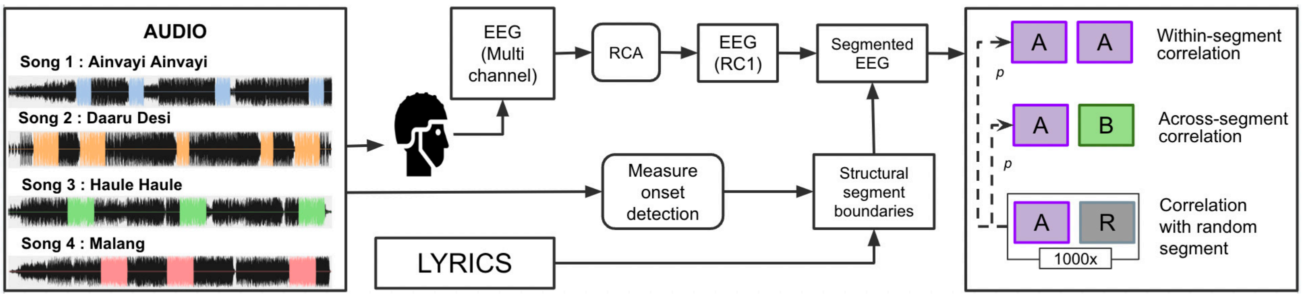

Figure 1. Analysis overview. The NMED-H dataset contains EEG responses recorded while 48 participants listened to four full-length Bollywood songs. We used RCA to compute a spatial EEG component in which ISC was maximized, and used the stimulus audio and lyrics to identify chorus segmentation boundaries to further epoch the EEG. For each song, correlations were performed within and across choruses, as well as between choruses and segments epoched at random.

Desi", "Haule Haule", and "Malang". Each song is around 4 min 30 sec in length and contains between 3 and 5 choruses as illustrated in color in Fig. 1. The stimuli were assumed to be new to the participants, who did not understand the Hindi-dialect song lyrics. We used the preprocessed, 125-channel EEG data sampled at 125 Hz with average reference; each song contained 24 trials from 12 unique participants (48 participants total) as each participant had listened to their assigned song twice.

3.2 EEG Analyses

To analyze the EEG, we followed an established procedure of spatial filtering followed by correlation calculations (Fig. 1). We used a publicly available RCA implementation [1] to compute a single spatial filter across all four songs. We computed RCA across entire song durations and not just chorus segments, as our permutation testing procedure involved segments sampled from throughout each song (see § 3.3). We then analyzed the vectorized form of single EEG trials from only the maximally reliable component RC1, as previous studies have shown that that component explains most of the ISC in EEG responses to music [30, 31]. Thus, the response data for each song was a time-by-trial matrix, with 24 trials from 12 participants for each song and a variable number of time samples per song.

To identify and segment song choruses, we first identified structural segment boundaries at the measure level using lyrics. [2] Next, we used a publicly available beat-tracking algorithm [38] to identify audio sample indices of the boundaries and converted those time stamps to the sampling rate of the EEG to segment the EEG accordingly.

Correlations were performed on a per-song basis, in two broad categories. *Within-chorus correlations* involved pairwise correlations among response trials from a single chorus instance, producing a symmetric matrix whose diagonal (being 1) was excluded from further analysis. *Across-chorus correlations* involved the cross-correlation of two matrices, each representing a different chorus instance. These correlations produced asymmetric matrices, since no response vector was ever correlated with itself.

[1] https://github.com/dmochow/rca
[2] https://gaana.com/, https://www.jiosaavn.com/

Each correlation also involved both *intra*-subject correlations (IaSC) of non-identical trials from the same participant and *inter*-subject correlations (ISC) of trials from different participants. As illustrated in Fig. 2, with 24 trials per song comprising two listens from each of 12 participants, within-chorus correlations produced for each participant one IaSC value (first listen and second listen) and 22 ISC values, excluding the diagonal. Across-chorus correlations produced for each participant four IaSC values (reflecting two distinct chorus instances $\times$ two distinct listens) and 88 ISC values. For each calculation, we computed mean correlations at the participant as well as the group level.

3.3 Statistical Analyses

We assessed statistical significance over distributions of per-participant results ($N = 12$). For RQ1 we used permutation testing: Each analysis was performed over 1000 pairs of segments of the same length as the true chorus segments, but with one segment epoched from a random start time in the song. The 1000 results served as the null distribution against which we compared the true result to compute the p-value. For RQ2 and RQ3 we used nonparametric Wilcoxon signed-rank tests to account for variable standard deviations of the sampling distributions caused by the discrepancy in the number of samples in each group (i.e., IaSC versus ISC; within- versus across-chorus). We performed one-sided tests in accordance with our expected results (RQ2 H_1: $within > across$; RQ3 H_1: $I_aSC > ISC$). We corrected for multiple comparisons using False Discovery Rate [39] on a per-song basis for RQ1 and RQ3 and on a per-song, per-condition basis for RQ2. We report statistically significant results ('***', '**') and also indicate but do not summarize marginally significant results ('*') for this first exploratory analysis.

4. RESULTS

4.1 Individual Correlations

We correlated vectors of spatially filtered, single-trial EEG on a per-song basis, both among responses to single choruses as well as across pairs of different choruses. The re-

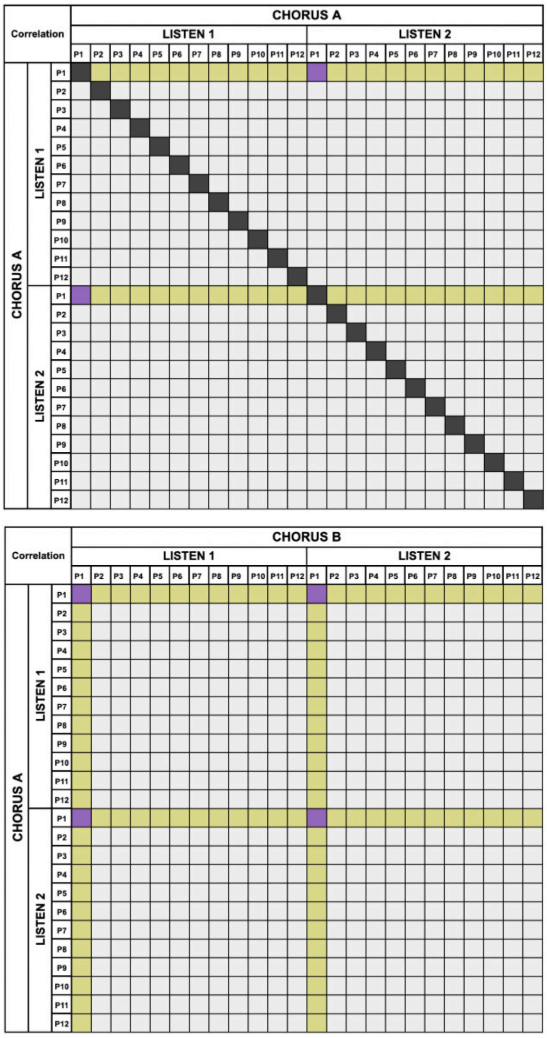

Figure 2. Illustration of IaSC and ISC matrix elements for Participant 1 of 12; each participant heard their assigned stimulus twice. **Top**: Within-chorus correlation produces a symmetric 24×24 matrix. The same IaSC correlation appears twice (purple), along with 22 unique ISC correlations (yellow). **Bottom**: Across-chorus correlation produces an asymmetric matrix with four unique IaSC correlations (2 choruses × 2 listens) and 88 unique ISC correlations.

sulting correlation matrices could then be partitioned into correlations from the same participant (IaSC) and different participants (ISC). Results are visualized in Fig. 3 and provided numerically in Tab. 1. After multiple comparison correction, 10 of 15 within-chorus IaSC and 2 of 22 across-chorus IaSC were statistically significant ('**' or '***'). For ISC, 14 of 15 within-chorus calculations and 12 of 22 across-chorus correlations were significant. IaSC distributions tended to have larger variance than ISC distributions, both at the participant level for single analyses (Fig. 3) and across the group means (Tab. 1).

4.2 Within- versus Across-Section Correlations

We assessed whether within-chorus correlations—involving identical musical content and context—were higher than across-chorus correlations, which are structurally similar but not identical. Tab. 2 summarizes the statistical significance of each comparison. After correcting for multiple comparisons, within-chorus IaSC was found to exceed across-chorus IaSC 7 times, while within-chorus ISC was higher than across-chorus ISC 4 times. Significant (and marginally significant) results most often implicated the first chorus of a song.

4.3 Intra- versus Inter-Subject Correlation

For our last analysis, we assessed whether IaSC—being computed from the same listener's data—would exceed ISC. Contrary to our expectations, one-sided Wilcoxon signed-rank tests revealed that after multiple comparison correction, IaSC did not exceed ISC for any within- or across-chorus correlation.

5. DISCUSSION

MSA has leveraged various representations—e.g., audio, lyrics, human annotations—to model human perception of musical structure. In this study we have answered the call for new forms of human response data to inform this task [1] and explored perception of repeated structure segments using brain data. Specifically, we assessed whether EEG responses to repeating choruses of four Bollywood songs were significantly correlated.

We found that EEG responses within and across choruses of a song were often significantly correlated, particularly for ISC. While small, these ISCs are on par with those reported in previous auditory EEG studies [30,31,40]. Correlating across choruses contrasts with past ISC research, which considered correlation only among responses to a single stimulus. That precedent may be due to those studies using predominantly narrative stimuli, such as movies or speeches, which generally do not include repeated segments. But for music, repetition is often integral to structure, from brief melodic motifs to large-scale elements [41]. The present use of ISC to assess music structure similarity is also a departure from its previous application to index brain states of attention and "engagement" in relation to attributes of stimuli (e.g., narrative tension, temporal coherence) [4, 30, 31] or participants (e.g., trained versus untrained musicians) [29]. Future research could consider data from spatial components beyond RC1 and further explore relationships between EEG correlation, music structure, and repetition to index both content similarity and listener engagement with repeated content.

We found that within-chorus correlation occasionally but not consistently exceeded across-chorus correlation; future research is needed to elucidate the role of acoustical or contextual differences across chorus instances in this result. Notably, within-chorus correlation most often exceeded across-chorus correlation in a song's first chorus. Past studies have shown that EEG-ISC often drops upon repeated exposures to full stimuli [4, 29, 30], and music-discovery engagement has been shown to be highest for first choruses compared to subsequent instances [42]. While this might lead one to expect higher ISC during

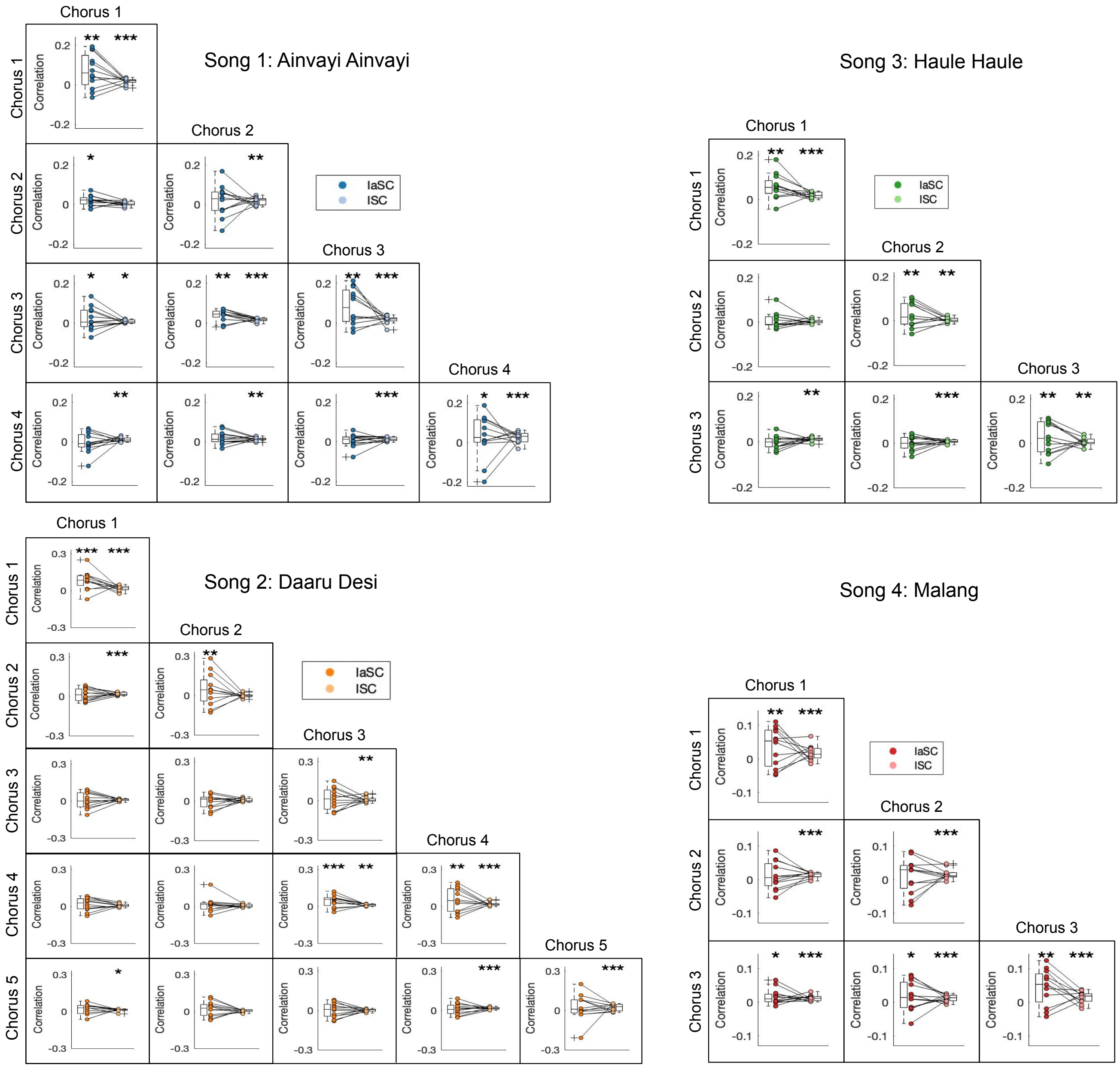

Figure 3. EEG correlations within and across choruses of four Bollywood songs. Each plot shows a distribution of intra- (IaSC) and inter- (ISC) subject correlation values across the 12 participants assigned to that song. Statistical significance of each correlation is denoted as p = 0 *** 0.01 ** 0.05 * 0.10 for FDR-corrected p-values.

the first chorus, current results do not suggest that within-chorus correlation drops as a song progresses. However, it may be that listeners have a unique perceptual experience of first choruses relative to other choruses.

Our expectation that IaSC would exceed ISC was not supported by the data. The large variance of IaSC relative to ISC, and greater number of significant ISC results despite lower group means, suggests that ISC ultimately provided a more stable estimate of neural correlation. Whether this is due to IaSC comprising fewer correlations, or an advantage of correlating across a heterogeneous sample of listeners, can be further investigated to inform future study designs.

This study contributes a first step toward using EEG data for MSA. While we focused on establishing similarity of neural responses among pre-identified repeating segments, and not detection of repeated segments or segment boundaries, our findings lay a foundation for multiple avenues of future work. For instance, a multimodal MSA framework could incorporate EEG measures of similarity alongside music content representations and human annotations. Other EEG-ISC analysis configurations may also prove useful for MSA: For instance, Dauer et al.'s finding that ISC computed over short time windows peaked during salient musical events including structural segment boundaries [31] is worth exploring further. Returning, too, to established connections between ISC and engagement, using ISC to identify highly engaging portions of songs could inform audio thumbnailing. Finally, while the present work leveraged an existing dataset, future studies could be de-

		IaSC					ISC				
		C1	C2	C3	C4	C5	C1	C2	C3	C4	C5
Song 1	C1	0.069**					0.017***				
	C2	0.020*	0.016				0.004	0.016**			
	C3	0.021*	0.041**	0.083**			0.008*	0.017***	0.017***		
	C4	-0.006	0.019	0.005	0.028*		0.010**	0.014**	0.013***	0.026***	
Song 2	C1	0.082***					0.019***				
	C2	0.013	0.046**				0.019***	0.001			
	C3	0.004	-0.003	0.013			0.010	0.007	0.010**		
	C4	0.018	0.014	0.044***	0.053**		0.010	0.006	0.011**	0.020***	
	C5	0.015	0.021	0.001	0.010	0.024	0.013*	0.006	0.001	0.016***	0.023***
Song 3	C1	0.059**					0.020***				
	C2	0.007	0.030**				0.003	0.005**			
	C3	-0.003	0.000	0.026**			0.010**	0.008***	0.007**		
Song 4	C1	0.036**					0.017***				
	C2	0.012	0.012				0.015***	0.017***			
	C3	0.017*	0.019*	0.045**			0.013***	0.012***	0.014***		

Table 1. Intra- and inter-subject correlation coefficients within and across choruses of four Bollywood songs. Statistical significance of correlations (FDR-corrected p-values) is denoted as p = 0 *** 0.01 ** 0.05 * 0.10.

		IaSC					ISC				
		C1	C2	C3	C4	C5	C1	C2	C3	C4	C5
Song 1	C1	-	*	*	*		-	**	**	*	
	C2	ns	-	ns	ns		ns	-	ns	ns	
	C3	ns	ns	-	**		ns	ns	-	ns	
	C4	ns	ns	ns	-		*	*	ns	-	
Song 2	C1	-	**	**	**	**	-	ns	ns	ns	ns
	C2	ns	-	ns	ns	ns	ns	-	ns	ns	ns
	C3	ns	ns	-	ns	ns	ns	ns	-	ns	ns
	C4	ns	ns	ns	-	ns	ns	ns	ns	-	ns
	C5	ns	ns	ns	ns	-	ns	*	*	ns	-
Song 3	C1	-	***	***			-	***	**		
	C2	ns	-	*			ns	-	ns		
	C3	ns	ns	-			ns	ns	-		
Song 4	C1	-	ns	ns			-	ns	ns		
	C2	ns	-	ns			ns	-	ns		
	C3	ns	ns	-			ns	ns	-		

Table 2. Results of one-sided Wilcoxon signed-rank tests assessing whether within-chorus correlation exceeds across-chorus correlation. Statistical significance of correlations (FDR-corrected p-values) is denoted as p = 0 *** 0.01 ** 0.05 * 0.10; ‘ns’ denotes non-significance.

signed to address specific MSA questions with newly collected EEG data. In all, we do not propose that EEG should or could replace existing data modalities for MSA, but rather highlight potential insights from EEG that may complement other existing approaches and inputs.

5.1 Limitations

We acknowledge limitations of this work. First, while we report multiple significant results, they do not imply generalizability: The correlations are small, and our findings—while promising—are not conclusive across all calculations. Next, we chose NMED-H as a ready-to-use EEG dataset of responses to popular songs containing repeated choruses. But the small stimulus set of four songs also hinders generalization, and future confirmatory studies should utilize a larger song set. We note that the original design of NMED-H specified that participants not be familiar with the songs or the language of their lyrics [35]. This too may limit generalizability, as more familiar or lyrically understandable songs may result in different EEG correlations.

Another main limitation is that while the song choruses crucially elicited the EEG data, they were only treated as repeating segments, and we did not consider nuances of placement or content attributes of individual choruses. Yet such features are known to impact perceptual and neural responses to choruses [26]. Thus, future research should consider finer-grained characterizations of music segments treated as structurally similar. One concrete next step could involve cross-modal comparisons of music similarity—for instance, whether similarity measures derived from audio, lyrics, or human annotations predict neural similarity.

Lastly, we trained RCA once over all available trials. Future work should incorporate cross-validation—iteratively optimizing the RCA spatial filter on training data and then applying it to holdout test trials—into the analysis pipeline to avoid overfitting.

6. CONCLUSION

MSA is an MIR topic with rich applications in audio thumbnailing, motif-finding, music summarization, music recommendation, and automatic music generation. Aiming to expand the scope of data modalities that may inform this task, we have contributed a first look at structural repetition using brain data. We used a publicly available EEG dataset and analyzed single-trial responses to choruses from four Bollywood pop songs by computing intra- and inter-subject correlations within and across choruses. We find that neural responses do often synchronize to a significant extent, which suggests that similarity among repeated choruses may translate to neural similarity. These findings motivate future studies of music similarity perception and highlight EEG data as a promising input to multimodal MSA systems.

7. ACKNOWLEDGMENTS

The authors thank Jacek Dmochowski for helpful advice on the statistical analyses.

8. REFERENCES

[1] O. Nieto, G. J. Mysore, C.-i. Wang, J. B. Smith, J. Schlüter, T. Grill, and B. McFee, "Audio-based music structure analysis: Current trends, open challenges, and applications," *Transactions of the International Society for Music Information Retrieval*, vol. 3, no. 1, 2020.

[2] J. Van Balen, J. A. Burgoyne, F. Wiering, R. C. Veltkamp *et al.*, "An analysis of chorus features in popular song," in *Proceedings of the 14th Society of Music Information Retrieval Conference*, 2013.

[3] B. Kaneshiro and J. P. Dmochowski, "Neuroimaging methods for music information retrieval: Current findings and future prospects," in *Proceedings of the 16th International Society for Music Information Retrieval Conference*, 2015, pp. 538–544.

[4] J. P. Dmochowski, P. Sajda, J. Dias, and L. Parra, "Correlated components of ongoing EEG point to emotionally laden attention—a possible marker of engagement?" *Frontiers in Human Neuroscience*, vol. 6, 2012.

[5] R. B. Dannenberg and M. Goto, *Music Structure Analysis from Acoustic Signals*. Springer, 2008, pp. 305–331.

[6] M. Goto, "A chorus-section detecting method for musical audio signals," in *2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03).*, vol. 5. IEEE, 2003, pp. V–437.

[7] J. Paulus and A. Klapuri, "Labelling the structural parts of a music piece with Markov models," in *Computer Music Modeling and Retrieval. Genesis of Meaning in Sound and Music: 5th International Symposium, CMMR*. Springer, 2009, pp. 166–176.

[8] A. Eronen and F. Tampere, "Chorus detection with combined use of MFCC and chroma features and image processing filters," in *Proc. of 10th International Conference on Digital Audio Effects*, 2007, pp. 229–236.

[9] B. McFee, O. Nieto, M. M. Farbood, and J. P. Bello, "Evaluating hierarchical structure in music annotations," *Frontiers in psychology*, vol. 8, p. 1337, 2017.

[10] R. S. Schaefer, J. Farquhar, Y. Blokland, M. Sadakata, and P. Desain, "Name that tune: Decoding music from the listening brain," *NeuroImage*, vol. 56, no. 2, pp. 843–849, 2011.

[11] S. Stober, D. J. Cameron, and J. A. Grahn, "Classifying EEG recordings of rhythm perception." in *Proceedings of the 15th International Society for Music Information Retrieval Conference*, 2014, pp. 649–654.

[12] M. S. Treder, H. Purwins, D. Miklody, I. Sturm, and B. Blankertz, "Decoding auditory attention to instruments in polyphonic music using single-trial EEG classification," *Journal of neural engineering*, vol. 11, no. 2, p. 026009, 2014.

[13] G. Cantisani, S. Essid, and G. Richard, "EEG-based decoding of auditory attention to a target instrument in polyphonic music," in *2019 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*. IEEE, 2019, pp. 80–84.

[14] G. Cantisani, G. Trégoat, S. Essid, and G. Richard, "MAD-EEG: An EEG dataset for decoding auditory attention to a target instrument in polyphonic music," in *Speech, Music and Mind (SMM), Satellite Workshop of Interspeech 2019*, 2019.

[15] G. Cantisani, S. Essid, and G. Richard, "Neuro-steered music source separation with EEG-based auditory attention decoding and contrastive-NMF," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2021, pp. 36–40.

[16] S. Stober, T. Prätzlich, and M. Müller, "Brain beats: Tempo extraction from EEG data," in *Proceedings of the 17th International Society for Music Information Retrieval Conference*, 2016, pp. 276–282.

[17] M.-S. Kim, G. Y. Lee, and H.-G. Kim, "Multi-channel EEG classification method according to music tempo stimuli using 3D convolutional bidirectional gated recurrent neural network," *The Journal of the Acoustical Society of Korea*, vol. 40, no. 3, pp. 228–233, 2021.

[18] G. Y. Lee, M.-S. Kim, and H.-G. Kim, "Extraction and classification of tempo stimuli from electroencephalography recordings using convolutional recurrent attention model," *ETRI Journal*, vol. 43, no. 6, pp. 1081–1092, 2021.

[19] A. Vinay, A. Lerch, and G. Leslie, "Mind the beat: Detecting audio onsets from EEG recordings of music listening," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021, pp. 231–235.

[20] A. Ofner and S. Stober, "Shared generative representation of auditory concepts and EEG to reconstruct perceived and imagined music," in *Proceedings of the 19th International Society for Music Information Retrieval Conference*. ISMIR, 2018, pp. 392–399.

[21] N. Gang, B. Kaneshiro, J. Berger, and J. P. Dmochowski, "Decoding neurally relevant musical features using Canonical Correlation Analysis," in *Proceedings*

of the 18th International Society for Music Information Retrieval Conference, 2017, pp. 131–138.

[22] J. R. Katthi and S. Ganapathy, "Deep multiway Canonical Correlation Analysis for multi-subject EEG normalization," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2021, pp. 1245–1249.

[23] C. Foster, D. Dharmaretnam, H. Xu, A. Fyshe, and G. Tzanetakis, "Decoding music in the human brain using EEG data," in *IEEE 20th International Workshop on Multimedia Signal Processing (MMSP)*, 2018, pp. 1–6.

[24] E. B. Abrams, E. Muñoz Vidal, C. Pelofi, and P. Ripollés, "Retrieving musical information from neural data: How cognitive features enrich acoustic ones," in *Proceedings of the 23rd International Society for Music Information Retrieval Conference*, 2022.

[25] A. Ofner and S. Stober, "Modeling perception with hierarchical prediction: Auditory segmentation with deep predictive coding locates candidate evoked potentials in EEG," in *Proceedings of the 21st International Society for Music Information Retrieval Conference*, 2020, pp. 566–573.

[26] S. Sangnark, P. Autthasan, P. Ponglertnapakorn, P. Chalekarn, T. Sudhawiyangkul, M. Trakulruangroj, S. Songsermsawad, R. Assabumrungrat, S. Amplod, K. Ounjai, and T. Wilaiprasitporn, "Revealing preference in popular music through familiarity and brain response," *IEEE Sensors Journal*, vol. 21, no. 13, pp. 14 931–14 940, 2021.

[27] U. Hasson, Y. Nir, I. Levy, G. Fuhrmann, and R. Malach, "Intersubject synchronization of cortical activity during natural vision," *Science*, vol. 303, no. 5664, pp. 1634–1640, 2004.

[28] J. P. Dmochowski, A. S. Greaves, and A. M. Norcia, "Maximally reliable spatial filtering of steady state visual evoked potentials," *NeuroImage*, vol. 109, pp. 63–72, 2015.

[29] J. Madsen, E. H. Margulis, R. Simchy-Gross, and L. C. Parra, "Music synchronizes brainwaves across listeners with strong effects of repetition, familiarity and training," *Scientific reports*, vol. 9, no. 1, pp. 1–8, 2019.

[30] B. Kaneshiro, D. T. Nguyen, A. M. Norcia, J. P. Dmochowski, and J. Berger, "Natural music evokes correlated EEG responses reflecting temporal structure and beat," *NeuroImage*, vol. 214, p. 116559, 2020.

[31] T. Dauer, D. T. Nguyen, N. Gang, J. P. Dmochowski, J. Berger, and B. Kaneshiro, "Inter-subject correlation while listening to minimalist music: A study of electrophysiological and behavioral responses to Steve Reich's Piano Phase," *Frontiers in Neuroscience*, vol. 15, 2021.

[32] S. Stober, A. Sternin, A. M. Owen, and J. A. Grahn, "Towards music imagery information retrieval: Introducing the OpenMIIR dataset of EEG recordings from music perception and imagination." in *Proceedings of the 16th International Society for Music Information Retrieval Conference*, 2015, pp. 763–769.

[33] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "DEAP: A database for emotion analysis using physiological signals," *IEEE transactions on affective computing*, vol. 3, no. 1, pp. 18–31, 2012.

[34] K. P. Miyapuram, N. Ahmad, P. Pandey, and J. D. Lomas, "Electroencephalography (EEG) dataset during naturalistic music listening comprising different genres with familiarity and enjoyment ratings," *Data in Brief*, vol. 45, p. 108663, 2022.

[35] B. Kaneshiro, D. T. Nguyen, J. P. Dmochowski, A. M. Norcia, and J. Berger, "Naturalistic music EEG dataset—Hindi (NMED-H)," in *Stanford Digital Repository*, 2016. [Online]. Available: http://purl.stanford.edu/sd922db3535

[36] S. Losorelli, D. T. Nguyen, J. P. Dmochowski, and B. Kaneshiro, "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*, 2017, pp. 339–346.

[37] B. Kaneshiro, D. T. Nguyen, J. P. Dmochowski, A. M. Norcia, and J. Berger, "Naturalistic music EEG dataset—Elgar (NMED-E)," in *Stanford Digital Repository*, 2021. [Online]. Available: https://purl.stanford.edu/pp371jh5722

[38] D. P. W. Ellis, "Beat tracking by dynamic programming," *Journal of New Music Research*, vol. 36, no. 1, pp. 51–60, 2007.

[39] Y. Benjamini and D. Yekutieli, "The control of the false discovery rate in multiple testing under dependency," *The Annals of Statistics*, vol. 29, no. 4, pp. 1165–1188, 2001.

[40] J. J. Ki, S. P. Kelly, and L. C. Parra, "Attention strongly modulates reliability of neural responses to naturalistic narrative stimuli," *Journal of Neuroscience*, vol. 36, no. 10, pp. 3092–3101, 2016.

[41] E. H. Margulis, *On repeat: How music plays the mind*. Oxford University Press, 2014.

[42] B. Kaneshiro, F. Ruan, C. W. Baker, and J. Berger, "Characterizing listener engagement with popular songs using large-scale music discovery data," *Frontiers in Psychology*, vol. 8, p. 416, 2017.