DECODING NEURALLY RELEVANT MUSICAL FEATURES USING CANONICAL CORRELATION ANALYSIS

Nick Gang Blair Kaneshiro Jonathan Berger
Center for Computer Research in Music and Acoustics
Stanford University
{ngang,blairbo,brg}@ccrma.stanford.edu

Jacek P. Dmochowski
Department of Biomedical Engineering
City College of New York
jdmochowski@ccny.cuny.edu

ABSTRACT

Music Information Retrieval (MIR) has been dominated by computational approaches. The possibility of leveraging neural systems via brain-computer interfaces is an alternative approach to annotating music. Here we test this idea by measuring correlations between musical features and brain responses in a statistically optimal fashion. Using an extensive dataset of electroencephalographic (EEG) responses to a variety of natural music stimuli, we employed Canonical Correlation Analysis to identify spatial EEG components that track temporal stimulus components. We found multiple statistically significant dimensions of stimulus-response correlation (SRC) for all songs studied. The temporal filters that maximize correlation with the neural response highlight harmonics and subharmonics of that song's beat frequency, with different harmonics emphasized by different components. The most stimulus-driven component of the EEG has an anatomically plausible, symmetric frontocentral topography that is preserved across stimuli. Our results suggest that different neural circuits encode different temporal hierarchies of natural music. Moreover, as techniques for decoding EEG advance, it may be possible to automatically label music via brain-computer interfaces that capture neural responses that are then translated into stimulus annotations.

1. INTRODUCTION

Computationally extracted audio features have been used in Music Information Retrieval (MIR) research to model the perceptual attributes of music. Music-specific features were first introduced by Tzanetakis & Cook for genre classification [43]. These and other features have been used in subsequent work for further study of genre [14, 21] and other music-tagging applications including emotion and mood classification [25, 35, 41] and artist identification [26].

By contrast, the music neuroscience community has historically focused primarily on experimental stimuli consisting of simple tones or short, synthesized instrumental melodies. This controlled paradigm allows for precise experimental manipulations, with the goal of investigating specific musical parameters; it also permits event-related averaging of responses over repeated trials. However, these stimuli lack the complexity and ecological validity of music that is consumed in real life, and preclude the study of global music processing [20].

In recent years, however, this field has increasingly utilized "naturalistic" music stimuli, including complete, real-world musical works. Here, the computationally extracted features developed for MIR research have found direct application as they provide objective, time-varying stimulus representations for which neural correlates can be investigated. To date, this approach has been successfully applied to a variety of brain imaging modalities including functional magnetic resonance imaging (fMRI) [1, 2, 40, 42], electroencephalography (EEG) [6, 22, 30, 38], and electrocorticography (ECoG) [31, 32, 37]. Both encoding (predicting neural activations from stimulus features) and decoding (predicting stimulus features from neural activations) approaches have been explored [2, 27, 40].

While neuroscience is not yet an established subfield of MIR, the approaches and insights of each field are arguably complementary [3, 16]. In the present study, we extend this interdisciplinary approach and investigate the relationship between time-varying features of naturalistic music and their EEG responses. We employ a hybrid encoding-decoding model to derive features and brain signals that maximally covary. The model temporally filters musical features while spatially filtering the EEG to learn a multidimensional mapping between stimulus and response, implemented here by Canonical Correlation Analysis. We uncover multiple statistically significant dimensions of stimulus-response correlation, with the first dimension showing a consistent EEG filter across different songs. Moreover, the temporal filters that maximize SRC emphasize harmonics and subharmonics of the beat frequency, with different harmonics selected by different dimensions of SRC. Our findings suggest that musical features can potentially be annotated by processing neural responses, opening up an entirely novel approach to MIR. Finally, all data and code will be made publicly available.

The remainder of the paper is organized as follows. In Section 2, we describe the EEG dataset, audio stimulus

Attribution: Nick Gang, Blair Kaneshiro, Jonathan Berger, and Jacek P. Dmochowski. "Decoding neurally relevant musical features using Canonical Correlation Analysis", 18th International Society for Music Information Retrieval Conference, Suzhou, China, 2017.

feature extraction, and analysis procedures. We present the results of our analyses in Section 3, and conclude with a discussion in Section 4.

2. METHODS

All analyses were performed using Matlab. [1]

2.1 EEG Dataset

Seeking ready-to-use EEG data reflecting natural music listening and for which we could obtain the stimuli, we used the publicly available NMED-H dataset [18]. This dataset contains EEG responses to intact and scrambled versions of full-length "Bollywood" songs, each approximately 4.5 minutes long. We used the responses to intact songs only, which comprise data from 48 unique participants (12 per song), who each heard their song twice—a total of 24 EEG trials per song. The data frames have been filtered and cleaned of ocular and noise artifacts, and contain recordings from 125 electrodes at a sampling rate of 125 Hz with average reference. Full details of data acquisition and preprocessing are given in Kaneshiro (2016) [15]. As the downloaded data contained missing values, we imputed missing data using a spatial average from neighboring electrodes before proceeding with analysis.

2.2 Stimulus Feature Extraction

The NMED-H documentation provides links to purchase the songs from iTunes, and instructions for converting them to the intact versions of the experimental stimuli [18]. After following those procedures, we extracted acoustical features using the MIR Toolbox, Version 1.5 [19]. We extracted the same collection of 20 short-term features that were used in a recent fMRI study by Alluri et al. [1]: Zero crossing rate, spectral centroid, high/low energy ratio, spectral spread, spectral rolloff, spectral entropy, spectral flatness, roughness, RMS energy, broadband spectral flux, and spectral flux for 10 octave-wide subbands. Features were extracted in 25-msec analysis windows with a 50% overlap between frames (standard parameters for short-term features [1, 43]), yielding a feature sampling frequency of 80 Hz. As in the Alluri study, we also orthogonalized the features using PCA, providing a lower-dimensional stimulus representation that contains contributions of all features under consideration [1]. We performed all subsequent analyses using PC1, as well as two individual features. RMS and spectral flux were chosen as they reflect amplitude envelope and timbre, respectively, and have been used in previous studies mapping music stimulus features to brain responses [1, 2, 30, 42].

As a reference for interpreting results, we extracted beat and tempo information from the stimulus audio files using a publicly available Matlab implementation [8]. [2] From the global tempo estimates, we computed frequencies relevant to processing hierarchical timescales in music, namely those corresponding to the beat (quarter note), as well as one fourth (whole note), half (half note), twice (eighth note), and four times (sixteenth note) the beat frequency. Previous studies have investigated contributions of beat frequencies to stimulus amplitude envelopes [28]; here we have taken a similar approach, visualizing low-frequency magnitude spectra of the three features used for analysis.

The audio waveforms, spectrograms, low-frequency magnitude spectra, and PC1 loadings of the four stimuli are shown in Fig. 1. By visual inspection, it is apparent that the four songs have different structures, and a variety of tempos. Furthermore, the feature FFTs show spectral peaks at both beat-relevant frequencies and other frequencies not occurring at multiples of the beat. Interestingly, PC1 loadings computed across the full set of 20 features are mostly consistent from song to song.

2.3 Canonical Correlation Analysis

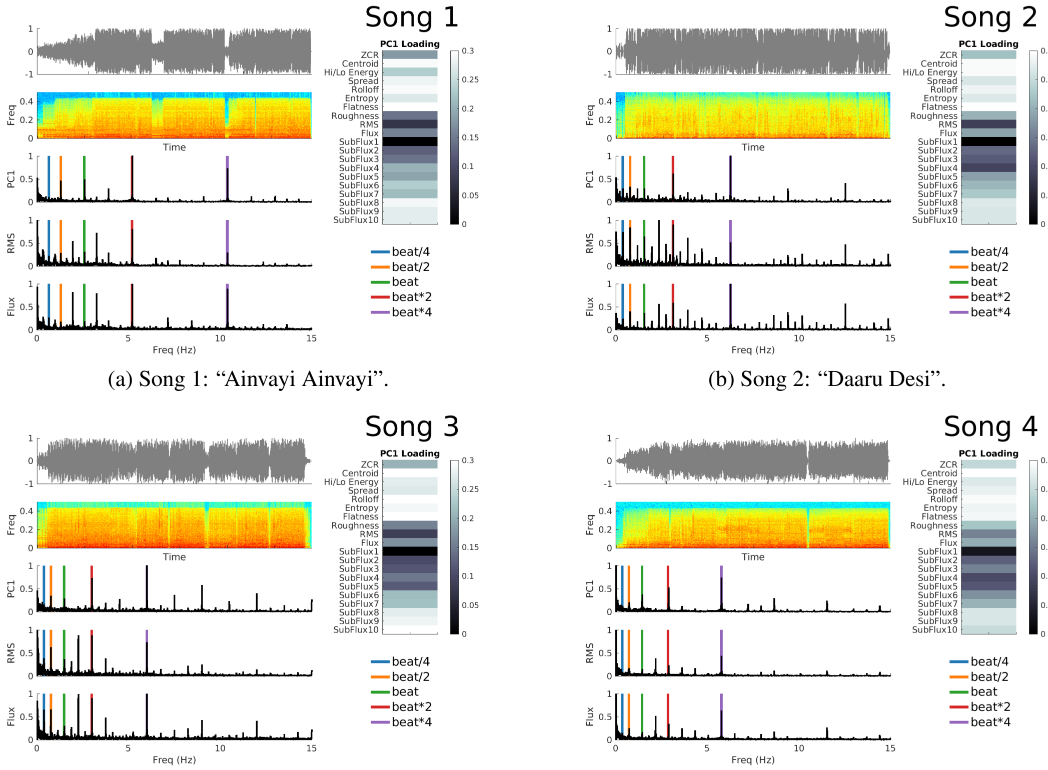
Canonical Correlation Analysis (CCA) involves projecting two data sets onto subspaces such that the projections are maximally correlated across time [9, 10, 13]. It has been used extensively in neuroscience, most recently as a technique for investigating links between visual stimuli and their EEG responses [7]. This approach may be thought of "hybrid encoding-decoding", in that the stimulus is temporally filtered (encoded) and the neural response spatially filtered (decoded), with the filtering optimized by CCA. The result is a multidimensional measure of the stimulus-response correlation (SRC), where each dimension emphasizes a different temporal component of the stimulus and a different spatial component of the EEG.

The inputs to the CCA are two matrices. For the present application, $X \in \mathbb{R}^{L \times T}$ is a convolution matrix of the stimulus feature where the row dimension spans time delays ("lags") and the column dimension spans time. In this construction, temporal filtering of the stimulus feature is achieved by multiplication with X. Matrix $Y \in \mathbb{R}^{D \times T}$ is the EEG data, where the row dimension spans electrodes and the column dimension spans time. CCA on X and Y produces a matrix of temporal filters $H \in \mathbb{R}^{L \times K}$ and a corresponding matrix of spatial filters $W \in \mathbb{R}^{D \times K}$ that extract temporal and spatial components from the stimulus and EEG, respectively, where K is the number of components. Therefore we obtain $U = H^T X$ and $V = W^T Y$, where U is a matrix of temporally filtered stimulus components, and V is a matrix of spatially filtered EEG components. The filters H and W are computed to maximize the correlation among corresponding rows of U and V (i.e., the components), under the constraint that the rows of U and V are temporally uncorrelated. The components are sorted in descending order of correlation, such that the first component pair (first rows of U and V) are most strongly correlated.

On a per-song basis, we pooled the data across trials to learn the model parameters. As the input sampling rates of the EEG and acoustical feature were not identical, we resampled the EEG to the sampling rate of the acoustical

[1] `https://www.mathworks.com/`

[2] `https://labrosa.ee.columbia.edu/projects/beattrack/`



(a) Song 1: “Ainvayi Ainvayi”.

(b) Song 2: “Daaru Desi”.

(c) Song 3: “Haule Haule”.

(d) Song 4: “Malang”.

Figure 1: Features of the songs used here as stimuli. From top left to bottom left in each pane are the waveform, spectrogram, and low frequency spectrum of each individual feature used (PC1, RMS energy, and spectral flux). On the right is the loading vector for the first PC computed across the 20 short-term features for that song.

features (80 Hz) beforehand. We performed separate CCA computations for each acoustical feature for each song, considering samplewise shifts up to 2 seconds in the construction of the feature input matrix X.

In sum, the CCA procedure outputs a delay-by-component stimulus filter matrix H, an electrodes-by-component response filter W, as well as the time samples-by-component filtered data outputs V and U. The Matlab code used to perform these analyses is made publicly available through GitHub. [3]

2.4 Visualizing the CCA Filters

While the columns of W provide the spatial filter weights, a “forward model” is recommended for visualizing component topographies on the scalp [12]. Thus, we used the EEG covariance matrix $R = YY^T$ to compute the forward-model projection $A = RW(W^TRW)^{-1}$ [29]. The columns of A represent the projection of the component onto the scalp and are visualized topographically.

For the temporal filters, we are interested primarily in their spectral characteristics, particularly at musically relevant (beat-related) frequencies. Therefore, we computed the FFT of each temporal filter and plotted its magnitude spectrum.

2.5 Stimulus-to-Response Correlations

The CCA procedure described above outputs U and V matrices containing the filtered data on a per-song, per-feature basis. We computed SRC for the first 5 components on a per-trial basis across the full duration of the trial. We report the mean correlation coefficient across trials, on a per-component, per-feature, per-song basis.

Due to autocorrelation characteristics of the stimulus and response data [37], we assessed statistical significance using a permutation test approach [39]. This was done by implementing the following procedure for each CCA computation performed above: First, we disrupted the temporal structure of individual trials of input EEG (while preserving aggregate spectral content) by phase scrambling the data from each electrode. Following that, the CCA and SRC computations were performed using the phase-scrambled EEG and intact acoustical feature as inputs. This procedure was repeated 500 times. We compared the SRC from intact data to the distribution of SRC across permutation iterations for computation of p-values. We corrected for multiple comparisons using False Discovery Rate (FDR) [4]. Reported statistical significance ($p < 0.05$) and marginal significance ($0.05 \leq p < 0.1$) reflect FDR correction.

[3] http://jd-lab.org/resources/

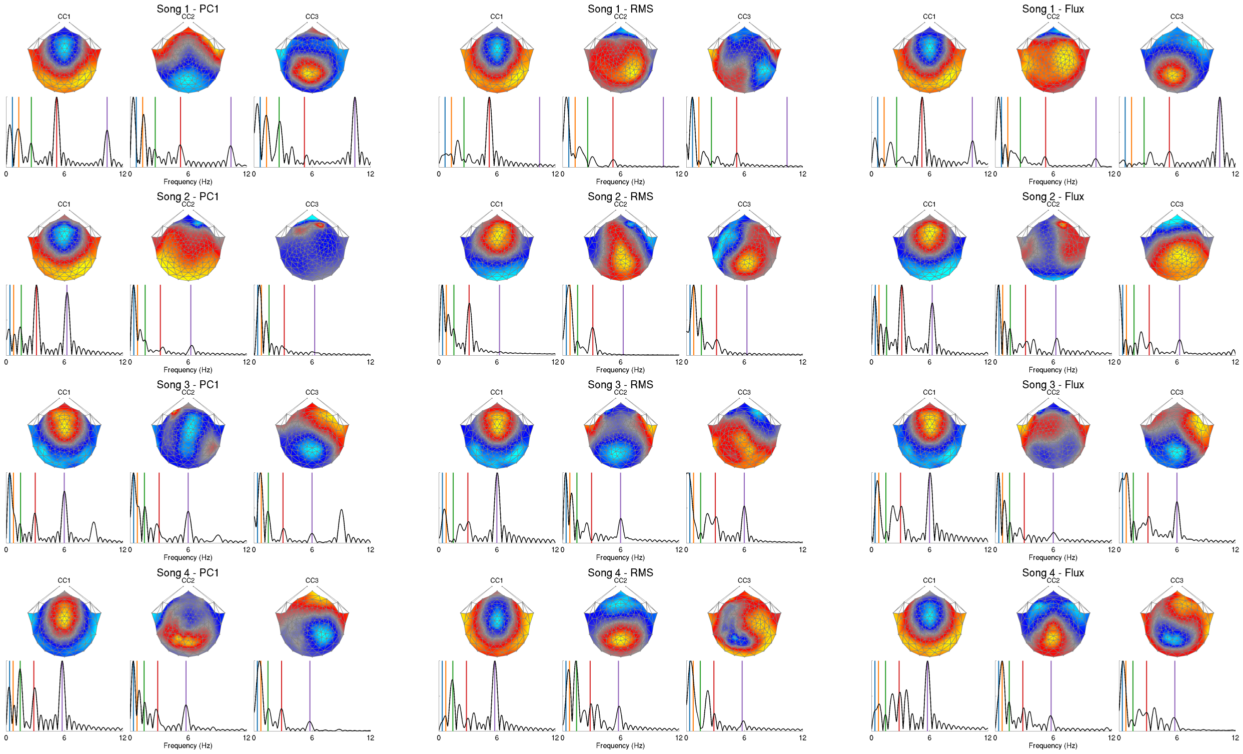


Figure 2: CCA filters. The spatial and temporal filters comprising each CCA pair are visualized for all songs and input features. Shown are the component topographies (spatial filters), as well as the frequency-domain representations of the temporal filters. The first 3 CCs are plotted. In each spectrum, vertical lines denote one fourth the beat frequency (blue), half the beat (orange), beat (green), twice the beat (red), and four times the beat (purple).

3. RESULTS

3.1 Spatial and Temporal Filters

We first probed the spatial topographies of the EEG components that best represented the musical features. Fig. 2 shows that the topography of component 1 is common across songs and features, up to a sign ambiguity inherent to CCA. The symmetric frontocentral topography of CC1 matches various past results involving spatial decomposition of brain responses during natural music listening [17,33,38]. Unlike the first CC, the second and third components tend to vary with the stimulus, but possess smooth and broad topographies consistent with the projections of cortical sources onto the scalp.

Interestingly, the temporal filters of each component are focused on harmonics and subharmonics of the song's beat frequency. In the case of CC1, the frequency responses of these filters tend to show peaks at higher beat-related frequencies (eighth and sixteenth notes). Subsequent CCs tend to show peaks at lower beat related frequencies (whole and half notes). Two exceptions to this are the filters for PC1 and spectral flux in Song 1. In both cases CC3 heavily emphasizes the sixteenth note frequency.

While some temporal filters within a single song and feature show similar frequency responses, these components can be differentiated by their phase. For example, Fig. 3 shows the time-domain representation of the temporal filters output for Song 1 RMS and Song 2 RMS. In both cases, the second and third filters emphasize whole note frequencies, but with a different phase. In general, all temporal filters show far more energy in beat-related frequencies than elsewhere.

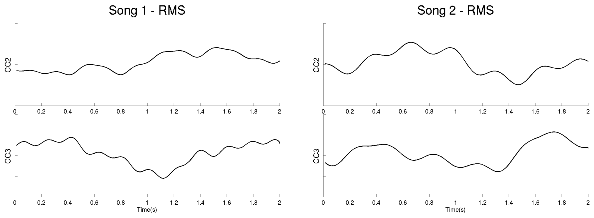


Figure 3: Time-domain examples of temporal filters with similar spectra but different phasing. Left: CC2 and CC3 for Song 1 RMS. Right: CC2 and CC3 for Song 2 RMS.

3.2 Stimulus-to-EEG Correlations

The results of our CCA procedure show multiple dimensions of significant correlation between the stimulus and brain response. As shown in Table 1, CC1 produces statistically significant SRC ($p < 0.05$ after FDR) for all songs and stimulus features. For the remaining components, the coefficients vary in strength across songs and features, but multiple dimensions of significance and marginal significance ($0.05 \leq p < 0.1$) are observed. We note that there is not a universal correspondence between correlation co-

	Comp.	PC1	Flux	RMS
Song 1	CC1	0.0652**	0.0690**	0.0630**
	CC2	0.0280**	0.0242**	0.0275*
	CC3	0.0212**	0.0179**	0.0153
	CC4	0.0177**	0.0179**	0.0123
	CC5	0.0135**	0.0102**	0.0115
Song 2	CC1	0.0522**	0.0573**	0.0524**
	CC2	0.0301**	0.0244**	0.0268**
	CC3	0.0183*	0.0177**	0.0203**
	CC4	0.0158**	0.0109**	0.0137*
	CC5	0.0119**	0.0062	0.0104**
Song 3	CC1	0.0460**	0.0530**	0.0437**
	CC2	0.0225	0.0306**	0.0254**
	CC3	0.0171*	0.0213**	0.0254**
	CC4	0.0139**	0.0119	0.0110
	CC5	0.0099*	0.0084	0.0074
Song 4	CC1	0.0536**	0.0511**	0.0475**
	CC2	0.0248	0.0324**	0.0261*
	CC3	0.0194*	0.0192**	0.0203**
	CC4	0.0170**	0.0178**	0.0135
	CC5	0.0114**	0.0102*	0.0089

Table 1: Multidimensional stimulus-response correlations captured by CCA. '**' denotes statistical significance ($p < 0.05$) and '*' denotes marginal significance ($0.05 \leq p < 0.1$) after correcting for FDR.

efficients and statistical significance. For example, in CC5 of Song 1, the correlation coefficient of $\rho = 0.0102$ for Flux is significant, while the slightly larger $\rho = 0.0115$ for RMS is not. This is due to the fact that separate permutation tests were performed, and surrogate EEG data generated, for each song and audio feature.

4. DISCUSSION

The technique outlined here provides a way to study music processing by direct comparison of an auditory stimulus and its corresponding brain response. Using CCA, matching spatial and temporal filters emerge that maximally correlate the stimulus and response in time. We found multiple dimensions of statistically significant correlation between stimulus and response. While the magnitudes of these correlations are small, the fact that they are not confined to a single dimension suggests that multiple brain areas process distinct portions of the stimulus. Such a multidimensional correlation could not be detected using sensor-space processing.

In past CCA studies using audio-visual stimuli [7], analysis of temporal filter resonances lacked clear relationships to the stimuli. However, the music studied here is organized by a hierarchy of beat- and measure-related periodicities, providing direct references with which to compare the temporal filter frequency responses. Here we found that the temporal filters that extract neurally relevant musical features are focused at harmonics of the beat frequency, independent of the song or feature.

Each CCA dimension emphasizes different brain sources (e.g., spatial topographies) and different combinations of harmonics. These results thus suggest that different temporal hierarchies of music are processed by distinct neural circuits. The topographical consistency of the strongest component, CC1, across all songs suggests a common mechanism for natural music listening. This consistency across songs is especially intriguing since the EEG data reflect disjoint sets of listeners for each song.

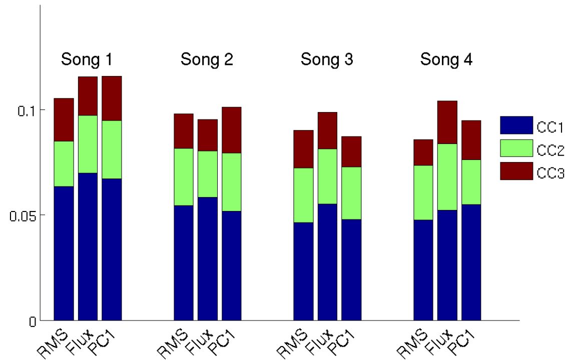


Figure 4: Total stimulus-response correlation for the first 3 CCs of each song and feature. The stacked bar graphs depict the proportion of stimuls-response correlation contributed by each CCA dimension.

From the analysis of SRC significance, PC1 does not appear to outperform other input features. In fact, spectral flux is the only feature that produces at least three significant components in each of the four songs. This may seem counterintuitive, but there is no guarantee that a feature explaining the most variance in the audio data will do the same for brain data. Indeed, the objective of CCA is to maximize covariance between the two data sets.

The correlations between musical features and EEG responses found here, while statistically significant, are fairly low (< 0.1). The low signal-to-noise ratio (SNR) of EEG severely limits the magnitudes of stimulus-response correlations, particularly with linear techniques as were used here. Moreover, the EEG recorded during music listening is driven mostly by sources unrelated to the auditory stimulus. The response to the stimulus comprises only a fractional component of the overall neural activity. Even with more sophisticated imaging modalities such as fMRI, correlation coefficients on the order of 0.1 are typically observed [11]. In order to increase the correlations between stimulus features and brain responses, nonlinear techniques such as deep neural networks could be employed in order to account for higher-order correlations and complex relationships not captured by linear CCA.

In research combining acoustical feature extraction and brain responses, it is important to consider the relative time scales on which stimuli and corresponding brain responses are sampled. Acoustical features are broadly separated into short- and long-term features. Past research has used the short-term features described above, as well as long-term "texture windows" with temporal resolution of around 1 Hz (e.g., 3-sec window with 33% overlap) [43]. Recording modalities for cortical responses can similarly be grouped by their temporal resolution. For example, EEG provides high temporal resolution (typically up to 1000 Hz for cortical responses), while fMRI offers a sampling frequency of only around 0.5 Hz [1]. Thus, in terms of time scales, EEG is amenable to short-term features and fMRI to long-term. Interestingly, however, many studies to date seem mismatched in this regard. There exist both fMRI studies utilizing short-term features [1,42] and EEG studies utilizing long-term features [6,22,44], meaning significant upsampling or downsampling was needed to compare stimulus features with responses. The present study is the first to our knowledge to non-invasively examine stimulus to response mapping in natural listening using exclusively matching timescales.

When choosing brain response recording modalities for this type of research, it is important to understand the tradeoffs in temporal and spatial resolution. Unlike hemodynamic signals of fMRI, the electrical signals recorded by EEG have been refracted through the skull and scalp; thus, observed topographies represent signals at specific electrodes, but not necessarily activations of specific underlying brain regions. ECoG methods solve this problem by placing electrodes directly on the cortex, but require invasive procedures and generally record from a smaller number of electrodes over a small region of the brain.

The present study correlated time-domain representations of both the acoustical features and EEG responses. The CCA approach could also be applied to transforms of either input. Past EEG and ECoG studies have examined time-frequency representations [6,22,31,32,37] and compared audio features with oscillatory band power in brain responses. Alternative stimulus input representations can also be considered. Time frequency transforms of the audio such as the Constant-Q Transform or other filterbank decompositions could be used as long- or short-term input features depending on the temporal resolution of interest. Using predetermined and hand-engineered features, as we did here, can also be limiting. The features used here are well represented in past research, but it could be beneficial for a system to learn the audio features themselves with the goal of improving the output of the optimization—for example with deep neural network approaches that have been applied to learn features for music tagging and signal processing systems [5,34].

Here we have chosen to average SRC coefficients for each song and feature across the full duration of the stimulus, producing a global correlation measure for each set of components. It is also possible to compute a time-varying measure of SRC and further investigate the musical events corresponding to moments of especially high or low SRC. Past research has even linked time-varying SRC to the attentional state of participants [7], pointing to application as a surrogate measure of listener attention. This approach could prove useful in an MIR context, providing a continuous, objective (brain-based) measure of attention to a real-world musical work.

While public access to naturalistic listening data remains limited, additional options exist. Given the limitations of the NMED-H dataset, it would be helpful to test this method on EEG datasets that reflect a wider range of musical genres [36] and tempos [23,24].

Future research may also consider differing stimuli across participants. Here, each CCA computation operated over concatenated EEG responses to a shared stimulus (e.g., all responses to Song 1). However, CCA has also been used to derive correlated components for unique perceptual experiences such as video game play [7]. MIR applications of this approach could involve pooling responses to different performances of the same song, or allowing participants to choose personal favorites. In addition, it will be interesting to investigate further the composition of the temporal stimulus filters, which for the present analyses are tightly coupled to beat frequencies, when songs of various tempos are analyzed together.

5. ACKNOWLEDGMENTS

This research was supported by the Wallenberg Network Initiative: Culture, Brain, Learning; the Roberta Bowman Denning Fund for Humanities and Technology; and the Patrick Suppes Gift Fund. The authors thank Anthony Norcia for helpful feedback on the analyses presented here.

6. REFERENCES

[1] V. Alluri, P. Toiviainen, I. P. Jääskeläinen, E. Glerean, M. Sams, and E. Brattico. Large-scale brain networks emerge from dynamic processing of musical timbre, key and rhythm. *NeuroImage*, 59(4):3677–3689, 2012.

[2] V. Alluri, P. Toiviainen, T. E. Lund, M. Wallentin, P. Vuust, A. K. Nandi, T. Ristaniemi, and E. Brattico. From Vivaldi to Beatles and back: predicting lateralized brain responses to music. *NeuroImage*, 83:627–636, 2013.

[3] J. J. Aucouturier and E. Bigand. Seven problems that keep MIR from attracting the interest of cognition and neuroscience. *Journal of Intelligent Information Systems*, 41(3):483–497, 2013.

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

[5] K. Choi, G. Fazekas, and M. Sandler. Automatic tagging using deep convolutional neural networks. In *ISMIR*, pages 805–811, 2016.

[6] F. Cong, V. Alluri, A. K. Nandi, P. Toiviainen, R. Fa, B. Abu-Jamous, L. Gong, B. G. W. Craenen, H. Poikonen, M. Huotilainen, and T. Ristaniemi. Linking brain responses to naturalistic music through analysis of ongoing EEG and stimulus features. *IEEE Transactions on Multimedia*, 15(5):1060–1069, 2013.

[7] J. P. Dmochowski, J. Ki, P. De Guzman, P. Sajda, and L. C. Parra. Extracting multidimensional stimulus-response correlations using hybrid encoding-decoding of neural activity. *NeuroImage*, 2017.

[8] D. P. W. Ellis. Beat tracking by dynamic programming. *Journal of New Music Research*, 36(1):51–60, 2007.

[9] H. R. Glahn. Canonical correlation and its relationship to discriminant analysis and multiple regression. *Journal of the Atmospheric Sciences*, 25(1):23–31, 1968.

[10] D. R. Hardoon, S. Szedmak, and J. Shawe-Taylor. Canonical correlation analysis: An overview with application to learning methods. *Neural Computation*, 16(12):2639–2664, 2004.

[11] Uri Hasson, Yuval Nir, Ifat Levy, Galit Fuhrmann, and Rafael Malach. Intersubject synchronization of cortical activity during natural vision. *science*, 303(5664):1634–1640, 2004.

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

[13] H. Hotelling. Relations between two sets of variates. *Biometrika*, 28(3/4):321–377, 1936.

[14] Y. F. Huang, S. M. Lin, H. Y. Wu, and Y. S. Li. Music genre classification based on local feature selection using a self-adaptive harmony search algorithm. *Data & Knowledge Engineering*, 92:60–76, 2014.

[15] B. Kaneshiro. *Toward an Objective Neurophysiological Measure of Musical Engagement*. PhD thesis, Stanford University, 2016.

[16] B. Kaneshiro and J. P. Dmochowski. Neuroimaging methods for music information retrieval: Current findings and future prospects. In *ISMIR*, pages 538–544, 2015.

[17] B. Kaneshiro, J. P. Dmochowski, A. M. Norcia, and J. Berger. Toward an objective measure of listener engagement with natural music using inter-subject EEG correlation. In *ICMPC13*, 2014.

[18] 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.

[19] O. Lartillot and P. Toiviainen. A Matlab toolbox for musical feature extraction from audio. In *DAFx*, pages 237–244, 2007.

[20] D. J. Levitin and V. Menon. Musical structure is processed in language areas of the brain: a possible role for brodmann area 47 in temporal coherence. *NeuroImage*, 20(4):2142–2152, 2003.

[21] T. Li, M. Ogihara, and Q. Li. A comparative study on content-based music genre classification. In *SIGIR*, pages 282–289, 2003.

[22] Y.-P. Lin, J.-R. Duann, W. Feng, J.-H. Chen, and T.-P. Jung. Revealing spatio-spectral electroencephalographic dynamics of musical mode and tempo perception by independent component analysis. *Journal of Neuroengineering and Rehabilitation*, 11(1):18, 2014.

[23] S. Losorelli, D. T. Nguyen, J. P. Dmochowski, and B. Kaneshiro. Naturalistic music EEG dataset—Tempo (NMED-T). In *Stanford Digital Repository*, 2017.

[24] 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 *ISMIR*, 2017.

[25] L. Lu, D. Liu, and H. J. Zhang. Automatic mood detection and tracking of music audio signals. *IEEE Transactions on Audio, Speech, and Language Processing*, 14(1):5–18, 2006.

[26] M. I. Mandel and D. P. W. Ellis. Song-level features and support vector machines for music classification. In *ISMIR*, pages 594–599, 2005.

[27] T. Naselaris, K. N. Kay, S. Nishimoto, and J. L. Gallant. Encoding and decoding in fMRI. *NeuroImage*, 56(2):400–410, 2011.

[28] S. Nozaradan, I. Peretz, and A. Mouraux. Selective neuronal entrainment to the beat and meter embedded in a musical rhythm. *The Journal of Neuroscience*, 32(49):17572–17581, 2012.

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

[30] H. Poikonen, V. Alluri, E. Brattico, O. Lartillot, M. Tervaniemi, and M. Huotilainen. Event-related brain responses while listening to entire pieces of music. *Neuroscience*, 312:58–73, 2016.

[31] C. Potes, P. Brunner, A. Gunduz, R. T. Knight, and G. Schalk. Spatial and temporal relationships of electrocorticographic alpha and gamma activity during auditory processing. *NeuroImage*, 97:188–195, 2014.

[32] C. Potes, A. Gunduz, P. Brunner, and G. Schalk. Dynamics of electrocorticographic (ECoG) activity in human temporal and frontal cortical areas during music listening. *NeuroImage*, 61(4):841–848, 2012.

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

[34] S. Sigtia and S. Dixon. Improved music feature learning with deep neural networks. In *ICASSP*, pages 6959–6963, 2014.

[35] Y. Song, S. Dixon, and M. Pearce. Evaluation of musical features for emotion classification. In *ISMIR*, pages 523–528, 2012.

[36] 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 *ISMIR*, 2015.

[37] I. Sturm, B. Blankertz, C. Potes, G. Schalk, and G. Curio. ECoG high gamma activity reveals distinct cortical representations of lyrics passages, harmonic and timbre-related changes in a rock song. *Frontiers in Human Neuroscience*, 8:798, 2014.

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

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

[40] P. Toiviainen, V. Alluri, E. Brattico, M. Wallentin, and P. Vuust. Capturing the musical brain with lasso: Dynamic decoding of musical features from fMRI data. *Neuroimage*, 88:170–180, 2014.

[41] K. Trohidis, G. Tsoumakas, G. Kalliris, and Ioannis P. Vlahavas. Multi-label classification of music into emotions. In *ISMIR*, pages 325–330, 2008.

[42] W. Trost, S. Frühholz, T. Cochrane, Y. Cojan, and P. Vuilleumier. Temporal dynamics of musical emotions examined through intersubject synchrony of brain activity. *Social cognitive and affective neuroscience*, page nsv060, 2015.

[43] G. Tzanetakis and P. Cook. Musical genre classification of audio signals. *IEEE Transactions on speech and audio processing*, 10(5):293–302, 2002.

[44] D. Wang, F. Cong, Q. Zhao, P. Toiviainen, A. K. Nandi, M. Huotilainen, T. Ristaniemi, and A. Cichocki. Exploiting ongoing EEG with multilinear partial least squares during free-listening to music. In *IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP)*, pages 1–6, 2016.