MUSIC IDENTIFICATION USING BRAIN RESPONSES TO INITIAL SNIPPETS

Pankaj Pandey[1]*, Gulshan Sharma*[2]*, Krishna. P. Miyapuram*[1]*, Ramanathan Subramanian*[3]*, Derek Lomas* [4]

[1] IIT Gandhinagar, [2] IIT Ropar, [3] U Canberra, [4] TU Delft

ABSTRACT

Naturalistic music typically contains *repetitive* musical patterns that are present throughout the song. These patterns form a signature, enabling effortless song recognition. We investigate whether neural responses corresponding to these repetitive patterns also serve as a signature, enabling recognition of later song segments on learning initial segments. We examine EEG encoding of naturalistic musical patterns employing the NMED-T and MUSIN-G datasets. Experiments reveal that (a) training machine learning classifiers on the initial 20s song segment enables accurate prediction of the song from the remaining segments; (b) β and γ band power spectra achieve optimal song classification, and (c) listener-specific EEG responses are observed for the same stimulus, characterizing individual differences in music perception.

Index Terms— Neural signatures, repetitive musical patterns, music perception, song identification

1. INTRODUCTION

Music perception triggers multiple neural processes, contributing to individual similarities and differences in music perception and appreciation. Activation of specific brain networks is dependent on idiosyncratic determinants including familiarity, musical training, musical engagement and sentiment. E.g., familiarity influences the cortical response to musical beat. Prior research has found that the neural response magnitude is limited for familiar as compared to unfamiliar, absurd music [1, 2]. Neural responses and cognitive performance of expert adult musical performers and non-experts differ significantly [3, 4, 5].

Repetition is a fundamental property of music, and repeating musical elements are typical of naturalistic music [6]. Naturalistic songs are easily identifiable because of these perceptible periodical patterns varying at a basic level. The repetitive property of music enables the human brain to easily correlate later song segments upon processing only the initial few seconds. However, depending upon individual idiosyncrasies, one might perceive the music uniquely. It is therefore tempting to understand the neural mechanism underlying musical stimulus processing. Over the last decade, several studies have sought to comprehend perceptual similarities and differences in audio-visual stimulus processing [7, 8, 9, 10].

Musical engagement can be comprehended through various ways such as listener states and mediums of listening. A recent study [11] shows varying levels of inter-subject correlation (ISC) across time to a shared real-world musical stimulus. Several studies on EEG-ISC suggest potential measures for explaining brain states related to engagement [12, 13]. There is also substantial literature discussing the impact of music on emotion-processing neural pathways [14, 15]. Overall, these findings motivate the need to examine the processing of inter-stimulus and inter-subject differences in music perception. Accordingly, we attempt to answer the following research questions in this work: (1) Are there significant correlations among an individual's neural responses across the length of a song? If strong correlations exist, it should be possible to recognize later portions of a song from neural (EEG) signals upon learning the initial responses. (2) Is the neural signature rooted in the initial segments preserved throughout the song? The extent of preservation would impact song recognition accuracy. (3) Is the neural signature associated with a song listener-specific or independent? This would determine whether a single model or multiple models would be required for EEG-based song recognition.

EEG data can be challenging to interpret based on simple visualisation because of their complex associations. However, there are multiple techniques that enable multivariate analysis and salient feature selection for a given task. EEG-based music research has shown promising results employing machine learning (ML) techniques; Stober and colleagues [16, 17] employ deep learning for song classification, and their primary objective is to maximize performance with convolutional neural networks. Foster and colleagues [18] correlate features extracted from the musical clips with corresponding EEG recordings. They employ the *Librosa* audio processing library to extract several features including Root Mean Square Error (RMSE), spectral roll-off, spectral centroid, chroma Short-Time Fourier Transform (STFT) and MFCCs. This is followed by pairwise correlation tests via representational similarity analysis and linear models. MFCCs and tempogram features are shown to correlate highly with EEG recordings, and a logistic regressor acheives accuracy of more than 20% to the chance level.

Differently, our study examines temporal consistencies in song-induced neural EEG signatures, and attempts intra- and inter-subject song classification utilizing only a few initial

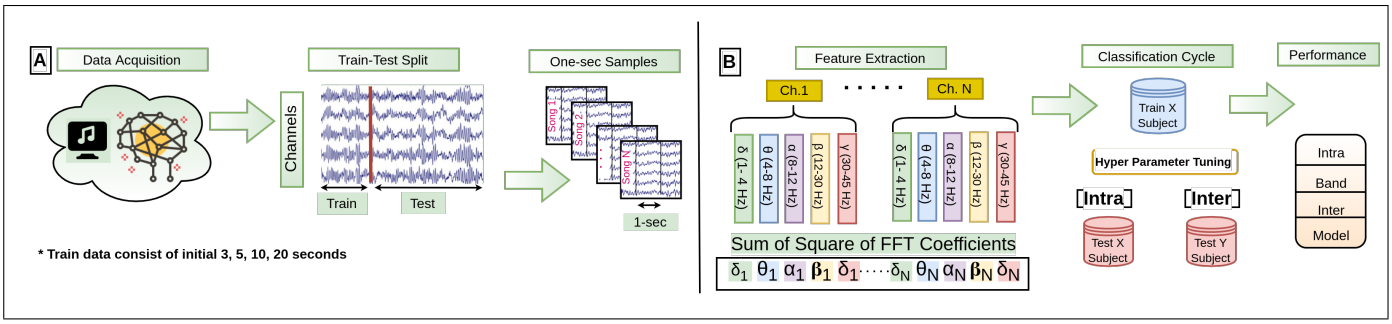

Fig. 1. Proposed Approach: (A) Segmentation of EEG Encodings. Train set contains initial segment of neural responses while the remaining segments constitute the test set. Segments are chunked into 1-second samples for analysis. (B) EEG signal decomposition and PSD feature computation. Followed by intra-subject and inter-subject song prediction models.

seconds of EEG recordings, and the importance of the aforementioned frequency bands to this end. *Intra-subject* refers to training and testing with the data of the same listener, whereas *inter-subject* refers to training a model on the data from one listener x, and testing on listener y's data. The main observations from our study are that: (a) Models trained on brain signals corresponding to initial few seconds of the song stimulus can reliably predict the EEG episodes for later song segments; (b) Different songs generate discriminative EEG signatures within the same listener (c) The β band contributes most to song recognition (d) Different brains generate varying signatures for an identical song, and (f) The random forest and LDA classifiers achieve maximum recognition performance.

2. EEG DATASETS (NMED-T AND MUSIN-G)

We analyzed two public naturalistic music EEG datasets. The Naturalistic Music EEG Dataset–Tempo (NMED-T) comprises EEG responses to ten songs whose lengths range between 4.5–5 minutes. Twenty participants (mean age 23) had their EEG recordings recorded with 128 electrodes. We used the preprocessed version comprising sampling rate to 125 Hz and 125 channels. Interested readers may refer to [19] about NMED-T acquisition and pre-processing.

Musing Listening-Genre EEG dataset (MUSIN-G) comprises EEG recordings for 12 musical stimuli with diverse genres [20]. Twenty participants were enrolled (mean age 23.5 years), aged 22-28 years. Song duration was two minutes. We preprocessed publicly available raw EEG data in MATLAB using EEGLAB [21]. Highpass linear FIR filtering at 0.3 Hz was performed, followed by the elimination of 50 Hz line noise using clean line method. A 250 Hz downsampling was applied. Upon removing bad channels, multiple Artifact Rejection Algorithm (MARA) was used to compute independent components and remove noisy components [22]. We interpolated removed channels via the spherical function of EEGLAB and an average reference was then applied. Cumulatively, this study examined EEG responses compiled from 40 subjects for 22 musical stimuli.

3. METHODS

3.1. Data Segmentation and Feature Extraction

Brain recordings are divided into train and test data. Train data contain only the initial seconds of music listening with window sizes of 3,5, 10, and 20s, and the remaining EEG recording used for testing. Data are segmented into 1s chunks to prepare input samples for feature extraction. These samples are of dimensions channels $\times$ time points (1s). E.g., the sample size is 128 $\times$ 250 (sampling rate) for MUSIN-G, and 125 $\times$ 125 for NMED-T. Oscillatory cortical activities in EEG time series are primarily present in the frequency bands ($\delta, \theta, \alpha, \beta$ and γ), power spectrum estimates of these frequency bands are computed across all channels. Time series signals are fed through Butterworth bandpass filters, and then Fast Fourier Transform (FFT) is performed on these time series for each epoch (1s EEG segment). Finally, sum of squared FFT coefficients are computed across each band to form a feature vector. Our approach is presented in Fig. 1.

3.2. Prediction Methods

We opted for classification methods, namely, Gaussian Naive Bayes, Linear Discriminant Analysis, linear Support Vector Machine and Random forest, for EEG-based song prediction. We performed 10 repetitions of 3-fold cross-validation on the data. As there is no class imbalance, *accuracy* is chosen as the performance metric. We trained models using windows of initial 3, 5, 10, and 20s EEG segments and evaluated classifiers on the later segments. In the MUSIN-G dataset, each participant listened to 12 songs. We evaluated the 12 song-classifier and report the average accuracy across participants, and likewise for NMED-T.

4. RESULTS

We examined intra and inter-subject song prediction performance, and the impact of the δ (1-4 Hz), θ (4-8 Hz), α (8-12 Hz), β (12-30 Hz), and γ (30-45 Hz) EEG frequency bands.

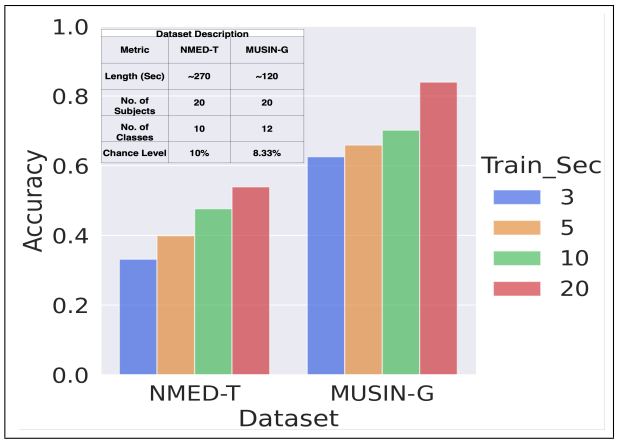

Fig. 2. Mean accuracy across participants for four training windows. Initial seconds are used for training and remaining for testing for each participant in NMED-T and MUSIN-G.

Band	RF	GNB	LDA	SVM	MLP
δ (1-4 Hz)	0.3	0.23	0.34	0.34	0.26
θ (4-8 Hz)	0.29	0.22	0.38	0.36	0.26
α (8-12 Hz)	0.25	0.18	0.37	0.34	0.24
β (12-30 Hz)	0.52	0.39	**0.61**	0.58	0.44
γ (30-40 Hz)	0.59	0.47	**0.65**	0.59	0.47
ALL-Bands	0.6	0.41	0.6	0.52	0.38

Table 1. NMED-T classifier performance with initial 40s for training. Bold font denotes two maximum accuracies.

Effect of varying window size: Initial EEG encodings of musical stimuli should preserve patterns to classify later EEG segments. In Fig. 2, we observed a significant improvement with an increase in training segment length. Interestingly, the initial 3s EEG recording can predict better than chance level. We obtained a maximum accuracy of 54%, and 84% utilizing initial EEG recordings of 20 sec in NMED-T, and MUSIN-G respectively. Lower accuracy for NMED-T could be due to the longer song length, and the initial 20s of EEG recording might not be sufficient for predicting later patterns. We further examined NMED-T using 40 seconds of training time and achieved maximum accuracy of 65% with the LDA classifier as shown in Table 1.

Same brain listens to different songs differently: We hypothesized that each song generates unique neural responses, and initial segments of these responses can predict the later segments. We rigorously tested our models and observed discriminating features within-subjects, with maximum accuracy of 97% and 71% in MUSIN-G and NMED-T, respectively. As shown in Fig.3, there is significant variability among subjects, which suggests that neural responses to different songs are different across individuals.

Significance of frequency bands for prediction: The β band showed the highest performance in MUSIN-G whereas γ performed slightly better for NMED-T. In the 1–12 Hz frequency range, θ band was dominant across all datasets. We also examined band combinations using three iterations of sequential forward feature selection [23]. Maximally discriminating bands were paired sequentially and evaluated

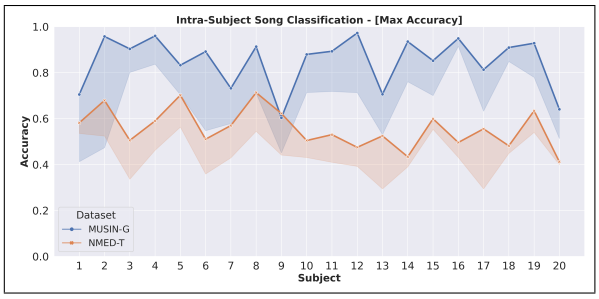

Fig. 3. Subject-wise performance on 20s of training data.

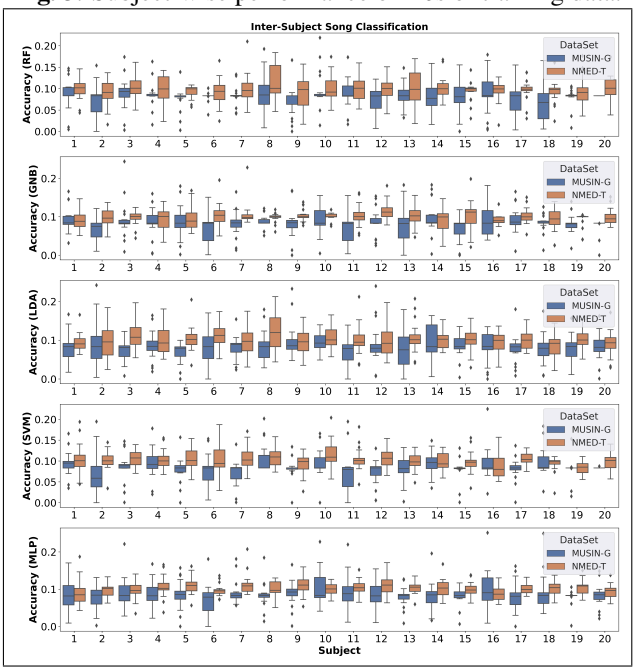

Fig. 4. Subject-independent song identification. Training on initial 20s of hold-out subject, and testing on other subjects.

(see Fig. 5). Role of β-band has been discussed widely in sensorimotor synchronization, which is associated with predicting new rhythms [24]. Shahin and colleagues find that music training promotes timbre-specific gamma-band activity [25]. γ band also correlates with perceptual and cognitive phenomena like template matching, feature binding, learning and memory formation.

Different brains listen to same song differently: To examine if neural signatures noted for the same song are subject-independent, we trained ML models on a single participant and evaluated them on the others. Compared with ML performance on individual data, the above approach produced significantly lower accuracies. We achieved a mean/max test accuracy of 12.9%/22.8% and 12.5%/24.5% on NMED-T and MUSIN-G respectively, as shown in Fig. 4. Inter-subject accuracies were higher in some cases, and very low in others. This finding implies that the same song generates varied neural patterns across listeners, and motivate the need to investigate EEG responses to musical features like *timbre* and *beat*, which may be more generic.

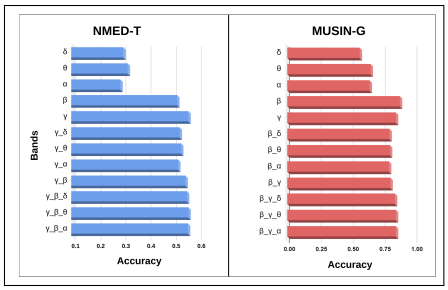

Fig. 5. Performance of frequency bands and combinations on 20s snippets with feature selection.

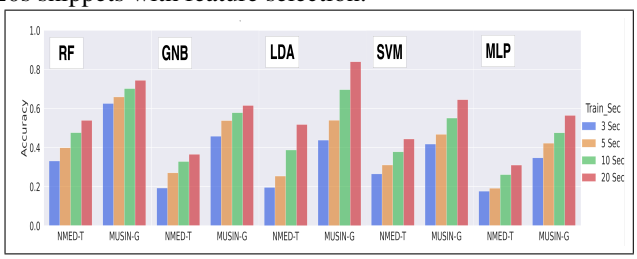

Fig. 6. ML-based intra-subject song prediction.

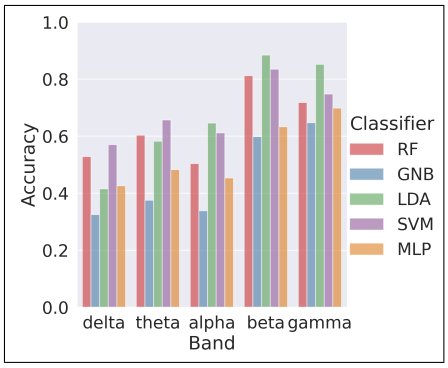

Fig. 7. (MUSIN-G) Prediction across frequency bands.

4.1. Classifier Performance

We observed that Random Forest and LDA outperformed other classifiers when utilizing all the frequency bands. As shown in Fig. 6, we observed at least 10% of increase in accuracy on LDA across varying window sizes in all the datasets . We also found that SVM performance increases when training on individual bands, as shown in Fig. 7. We achieved a maximum intra-subject accuracy of 65% using the γ band features in NMED-T, and accuracies of 88% for MUSIN-G using β band features.

5. CONCLUSION

Musically induced neural signals were examined in this research with three primary objectives: (1) EEG responses should capture repetitive characteristics of music that could be present throughout the length of the song, (2) An individual's brain signals should be discriminative of the different songs that are listened by an individual, and (3) Different individuals should generate different neural signatures corresponding to the same music. We carried out this research using two music datasets employing a variety of machine learning techniques. We find that small segments capturing initial brain responses enable sufficient learning of EEG signatures in the spectral domain. Higher frequency bands, namely β and γ neural oscillations provide the most discriminating features. For intra-subject song prediction, we achieve a maximum accuracy of 65% using γ features in NMED-T. The β band achieves 88% accuracy for MUSIN-G. Prediction accuracy drops significantly in inter-subject song classification, suggesting a weak correlation in brain responses among subjects. Given that studies on cognitive load assessment [26] and EEG-based fake-video detection [27] have shown subject-independent correlations between neural responses and stimulus patterns, future work would focus on identifying neural correlates underlying naturalistic musical perception irrespective of individual experiences.

6. REFERENCES

[1] Yuiko Kumagai, Mahnaz Arvaneh, and Toshihisa Tanaka, "Familiarity affects entrainment of eeg in music listening," *Frontiers in human neuroscience*, vol. 11, pp. 384, 2017.

[2] Benjamin Meltzer, Chagit S Reichenbach, Chananel Braiman, Nicholas D Schiff, AJ Hudspeth, and Tobias Reichenbach, "The steady-state response of the cerebral cortex to the beat of music reflects both the comprehension of music and attention," *Frontiers in human neuroscience*, vol. 9, pp. 436, 2015.

[3] Felipe Porflitt and Ricardo Rosas, "Core music elements: rhythmic, melodic and harmonic musicians show differences in cognitive performance (elementos básicos de la música: músicos rítmicos, melódicos y armónicos muestran diferencias de desempeño cognitivo)," *Studies in Psychology*, vol. 41, no. 3, pp. 532–562, 2020.

[4] Helmuth Petsche, K Linder, Peter Rappelsberger, and Gerold Gruber, "The eeg: An adequate method to concretize brain processes elicited by music," *Music Perception*, vol. 6, no. 2, pp. 133–159, 1988.

[5] Takako Fujioka, Bernhard Ross, Ryusuke Kakigi, Christo Pantev, and Laurel J Trainor, "One year of musical training affects development of auditory cortical-evoked fields in young children," *Brain*, vol. 129, no. 10, pp. 2593–2608, 2006.

[6] Zafar Rafii and Bryan Pardo, "A simple music/voice separation method based on the extraction of the repeating musical structure," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2011, pp. 221–224.

[7] Pankaj Pandey, Nashra Ahmad, Krishna Prasad Miyapuram, and Derek Lomas, "Predicting dominant beat frequency from brain responses while listening to music," in *2021 IEEE International Conference on Bioinformatics and Biomedicine*, 2021, pp. 3058–3064.

[8] Mojtaba Khomami Abadi, Ramanathan Subramanian, Seyed Mostafa Kia, Paolo Avesani, Ioannis Patras, and Nicu Sebe, "DECAF: Meg-based multimodal database for decoding affective physiological responses," *IEEE Transactions on Affective Computing*, vol. 6, no. 3, pp. 209–222, 2015.

[9] Ramanathan Subramanian, Julia Wache, Mojtaba Khomami Abadi, Radu L. Vieriu, Stefan Winkler, and Nicu Sebe, "Ascertain: Emotion and personality recognition using commercial sensors," *IEEE Transactions on Affective Computing*, vol. 9, no. 2, pp. 147–160, 2018.

[10] Dhananjay Sonawane, Krishna Prasad Miyapuram, Bharatesh Rs, and Derek J. Lomas, "Guess the music: Song identification from Electroencephalography response," in *8th ACM IKDD CODS and 26th COMAD*, 2021, p. 154–162.

[11] Blair Kaneshiro, Duc T Nguyen, Anthony Matthew Norcia, Jacek P Dmochowski, and Jonathan Berger, "Inter-subject EEG correlation reflects time-varying engagement with natural music," *bioRxiv*, 2021.

[12] Kat Agres, Dorien Herremans, Louis Bigo, and Darrell Conklin, "Harmonic structure predicts the enjoyment of uplifting trance music," *Frontiers in psychology*, vol. 7, pp. 1999, 2017.

[13] Blair Kaneshiro, Duc T Nguyen, Anthony M Norcia, Jacek P Dmochowski, and Jonathan Berger, "Natural music evokes correlated eeg responses reflecting temporal structure and beat," *NeuroImage*, vol. 214, pp. 116559, 2020.

[14] Joana Sa de Almeida, Lara Lordier, Benjamin Zollinger, Nicolas Kunz, Matteo Bastiani, Laura Gui, Alexandra Adam-Darque, Cristina Borradori-Tolsa, François Lazeyras, and Petra S Hüppi, "Music enhances structural maturation of emotional processing neural pathways in very preterm infants," *NeuroImage*, vol. 207, pp. 116391, 2020.

[15] Ian Daly, Asad Malik, Faustina Hwang, Etienne Roesch, James Weaver, Alexis Kirke, Duncan Williams, Eduardo Miranda, and Slawomir J Nasuto, "Neural correlates of emotional responses to music: an eeg study," *Neuroscience letters*, vol. 573, pp. 52–57, 2014.

[16] Sebastian Stober, Daniel J Cameron, and Jessica A Grahn, "Using convolutional neural networks to recognize rhythm stimuli from electroencephalography recordings," in *Advances in neural information processing systems*, 2014, pp. 1449–1457.

[17] Sebastian Stober, Avital Sternin, Adrian M Owen, and Jessica A Grahn, "Deep feature learning for eeg recordings," *arXiv preprint arXiv:1511.04306*, 2015.

[18] Chris Foster, Dhanush Dharmaretnam, Haoyan Xu, Alona Fyshe, and George Tzanetakis, "Decoding music in the human brain using eeg data," in *2018 IEEE 20th International Workshop on Multimedia Signal Processing (MMSP)*. IEEE, 2018, pp. 1–6.

[19] Steven Losorelli, Duc T Nguyen, Jacek P Dmochowski, and Blair Kaneshiro, "Nmed-t: A tempo-focused dataset of cortical and behavioral responses to naturalistic music.," in *ISMIR*, 2017, pp. 339–346.

[20] Krishna Prasad Miyapuram, Pankaj Pandey, Nashra Ahmad, Bharatesh R Shiraguppi, Esha Sharma, Prashant Lawhatre, Dhananjay Sonawane, and Derek Lomas, ""music listening- genre eeg dataset (musin-g)"," 2021.

[21] Arnaud Delorme and Scott Makeig, "Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis," *Journal of neuroscience methods*, vol. 134, no. 1, pp. 9–21, 2004.

[22] Irene Winkler, Stefan Haufe, and Michael Tangermann, "Automatic classification of artifactual ica-components for artifact removal in eeg signals," *Behavioral and Brain Functions*, vol. 7, no. 1, pp. 1–15, 2011.

[23] Sebastian Raschka, "http://rasbt.github.io/mlxtend/," .

[24] Vanessa Krause, Alfons Schnitzler, and Bettina Pollok, "Functional network interactions during sensorimotor synchronization in musicians and non-musicians," *Neuroimage*, vol. 52, no. 1, pp. 245–251, 2010.

[25] Antoine J Shahin, Larry E Roberts, Wilkin Chau, Laurel J Trainor, and Lee M Miller, "Music training leads to the development of timbre-specific gamma band activity," *Neuroimage*, vol. 41, no. 1, pp. 113–122, 2008.

[26] Maneesh Bilalpur, Mohan Kankanhalli, Stefan Winkler, and Ramanathan Subramanian, "Eeg-based evaluation of cognitive workload induced by acoustic parameters for data sonification," in *International Conference on Multimodal Interaction*, 2018, p. 315–323.

[27] Parul Gupta, Komal Chugh, Abhinav Dhall, and Ramanathan Subramanian, "The eyes know it: Fakeet-an eye-tracking database to understand deepfake perception," in *International Conference on Multimodal Interaction*, 2020, p. 519–527.