ENHANCING PREDICTIVE MODELS OF MUSIC FAMILIARITY WITH EEG: INSIGHTS FROM FANS AND NON-FANS OF K-POP GROUP NCT127

Seokbeom Park[1] **Hyunjae Kim**[1] **Kyung Myun Lee**[1,2]
[1] Graduate School of Culture Technology, KAIST, South Korea
[2] School of Digital Humanities and Computational Social Sciences, KAIST, South Korea
{fpfmxh, present, kmlee2}@kaist.ac.kr

ABSTRACT

Predicting a listener's experience of music based solely on audio features has its limitations due to the individual variability in responses to the same music. This study examines the effectiveness of electroencephalogram (EEG) in predicting the subjective experiences while listening to music, including arousal, valence, familiarity, and preference. We collected EEG data alongside subjective ratings of arousal, valence, familiarity, and preference from both fans (N=20) and non-fans (N=34) of the K-pop idol group, NCT127 to investigate response variability to the same NCT127 music. Our analysis focused on determining whether the inclusion of EEG alongside audio features could enhance the predictive power of linear mixed-effect models for these subjective ratings. Specifically, we employed stimulus-response correlation (SRC), a recent approach in neuroscience correlating stimulus features with EEG responses to the ecologically valid stimuli. The results showed that familiarity and preference was significantly higher in the fan group. Furthermore, the inclusion of SRC significantly enhanced the prediction of familiarity compared to models based solely on audio features. However, the impact of SRC on predictions of arousal and valence exhibited variation depending on the correlated audio features, with certain SRCs improving predictions while others diminished them. For preference, only a few SRCs negatively affected model performance. These results suggest that correlations of EEG responses and audio features can provide information of individual listeners' subjective responses, particularly in predicting familiarity.

1. INTRODUCTION

The neuroscience of music, employing neuroimaging methods, has revealed how the brain processes music through regions responsible for auditory, motor, and emotional functions, with recent approaches focusing on the brain's predictive processes [1, 2, 3, 4]. The convergence of music information retrieval (MIR) and neuroscience has gained significant traction in recent years [5, 6, 7]. For example, Rajagopalan and Kaneshiro have highlighted the potential of electroencephalogram (EEG) in the analysis of musical structure [8]. Furthermore, Ofner and Stober demonstrated the reconstruction of perceived and imagined music from EEG data [9]. These findings highlight the synergistic benefits of integrating MIR and neuroscience. In this paper, we aim to investigate how EEG can enhance the predictive model of subjective listening responses to music, given the individual variability in such experiences.

1.1 Predicting Subjective Music Listening Experience using Audio Features

Subjective music listening experience refers to the individual and unique responses that people have when they listen to music. It encompasses a wide range of aspects, including emotional reactions, preferences, familiarity, and overall enjoyment of the music. Subjective experience acknowledges that each listener's response to music is personal and may be influenced by various factors such as their musical background and cultural upbringing [10, 11, 12, 13].

Predicting listeners' subjective experiences of music through audio features has been a significant focus within MIR research. For example, Music Emotion Recognition (MER) aims to predict listeners' emotional responses using various techniques [14, 15, 16, 17]. Audio features, including tempo, rhythm, melody, and harmony, have been shown to correlate with listeners' emotional responses and preferences [18, 17]. However, the relationship between audio features and subjective experiences is complex, influenced by individual differences in musical background, culture, and personal taste [19, 20, 21]. Notably, emotional responses can significantly vary depending on individual differences [21, 22, 23, 24]. Thus, relying solely on audio features may not capture the full spectrum of music's impact on the listener, emphasizing the need for incorporating physiological measures such as EEG in understanding subjective music experiences [18].

1.2 Stimulus-response Correlation

The use of EEG offers a breakthrough in predicting subjective music listening experiences [25]. EEG provides real-time measures of brain activity, allowing direct observa-

 Attribution: S. Park, H. Kim, and K.M. Lee, "Enhancing predictive models of music familiarity with EEG: Insights from fans and non-fans of K-pop group NCT127", in *Proc. of the 25th Int. Society for Music Information Retrieval Conf.,* San Francisco, United States, 2024.

tion of neural responses to musical stimuli. Specifically, stimulus-response correlation (SRC), a recent method correlating stimulus features with EEG responses, enhances the ecological validity of studies by using real-world music stimuli and offers interpretable insights into the direct effects of stimulus features on the listener's experience [26, 27]. For example, SRC analysis revealed that neural responses are strongly correlated with specific task-relevant visual areas [28]. Additionally, SRC enabled the prediction of speech intelligibility [29]. Despite employing a different method to calculate the correlation between audio features and EEG responses, Weineck et al. found that neural response intensity increased with music familiarity [30]. Therefore, SRC is considered to be useful tool for predicting subjective music listening experiences.

1.3 Research Question

In this study, we aim to investigate the variability of subjective music listening experiences by comparing responses of fans and non-fans to K-pop idol music. Subsequently, we explore the effectiveness of SRC in predicting this individual response variability. To achieve this goal, we formulated the following research questions:

RQ 1: How do subjective music listening experiences, such as arousal, valence, familiarity, and preference, vary individually for the same music among fans and non-fans of K-pop idol music?

RQ 2: How does the inclusion of SRC alongside audio features affect the predictive power of models for arousal, valence, familiarity, and preference in music listening?

RQ 3: Does the effectiveness of SRC in predicting subjective experiences vary depending on the type of audio feature it is correlated with?

To address these questions, we conducted an experiment collecting EEG data and subjective ratings from both fans and non-fans of NCT127 as they listened to music by the group. Utilizing linear mixed-effects models, we analyzed the contribution of audio features and SRC in predicting subjective experiences, providing a comprehensive understanding of how these components interact to shape individual music listening experiences.

2. MATERIALS AND METHODS

2.1 Participants

We recruited 20 fans of NCT127 (mean age 24.8 years, 2 males) and 34 non-fans (mean age 26.1 years, 7 males). To participate in the experiment as part of the fan group, participants were required to meet at least one of the following conditions: they must have attended at least one event featuring NCT127, such as a concert or fan meeting, or they must own at least one piece of NCT127-related merchandise, such as an album, light stick, photocard, LP, or sheet music. This was verified through a photo submission process when applying for the experiment. All participants were Korean non-musicians. All participants had normal hearing and provided written informed consent before starting the experiment.

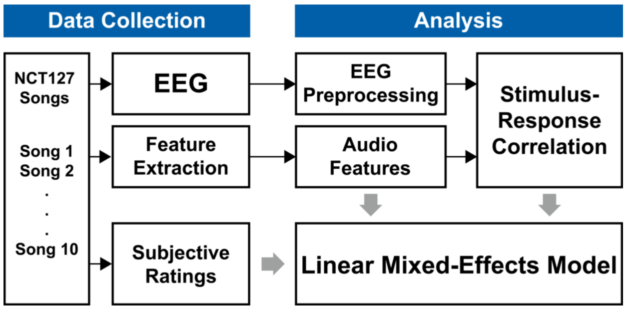

Figure 1: Schematic view of data collection and analysis.

2.2 Stimuli

The music used in the experiment consisted of the NCT127's top 10 songs based on the YouTube Music rankings as of December 26, 2023. The music was edited from the beginning to the end of the first chorus. The length of the edited audio varied between 60 to 92 seconds. Each audio was edited to began with a 0.5 second fade-in and ended with a 0.5 second fade-out. Then, volume normalization was applied to each channel before being exported. As a result, ten stereo audio files with a 44100Hz sampling rate and 16-bit depth were created for the stimuli.

2.3 Experiment

The EEG experiment was conducted using the Compumedics Neuroscan system. For EEG recordings, a Synamp RT 64-channel amplifier and a 64-channel Quik-Cap with sintered Ag/AgCl electrodes were used. The data collection was carried out through the Curry 8 acquisition software. EEG electrodes were placed in accordance with the international 10-20 system, and EEG data were collected at a sampling rate of 1000Hz across 64 channels.

The experiment was conducted using STIM2 software in a soundproof room to eliminate noise interference. Participants listened to each stimulus through insert earphone while focusing on a cross in the center of the monitor. Each stimulus was played once, and after listening to each, participants rated their arousal, valence, familiarity, and preference using a 7-point scale. Participants were able to proceed to the next stimulus after completing their ratings. There was a 5-second silence window before and after each stimulus, and the stimuli were played in a randomized order. An overall view of the data collection and analysis is presented in Figure 1.

2.4 Analysis

2.4.1 EEG Preprocessing

The preprocessing of EEG data was conducted using MATLAB with the EEGLAB toolbox [31]. From the 64 channels, the reference channels M1 and M2 were excluded. The EEG data underwent a 1-55 Hz bandpass FIR filter, followed by epoching for each stimulus. Subsequent steps included baseline removal and downsampling from 1000Hz to 125Hz. The data were re-referenced using the common average reference method, and all EEG data were merged by each participant. Independent Component Analysis (ICA) decomposition (using *runamica15* function) was performed to remove artifacts [32]. Artifactual components (eye, muscle, heart) were chosen by automated artifact IC classifier 'IClabel' and additional artifactual components were manually chosen [33]. Finally, the EEG data were epoched by each stimulus.

2.4.2 *Stimulus-response Correlation*

To calculate SRC, we applied a hybrid encoding-decoding technique, performing canonical correlation analysis to maximize the correlation between temporally filtered stimuli (audio) and spatially filtered neural responses (EEG). A detailed explanation of the method, including the computation of spatial and temporal response functions for each component, can be found in Dmochowski et al.'s paper [26].

For SRC calculations, stimulus features were extracted from each audio stimulus using MATLAB `mirtoolbox` [34]. From audio features that Lange and Frieler explored [18], only audio features permitting extraction in a time-by-feature value manner, thus enabling SRC calculation, were selected for investigation. This process resulted in extracting ten audio features: sound envelope, root mean square (RMS), spectral flux, zero-crossing rate, roughness, spectral entropy, spectral centroid, spectral spread, spectral rolloff, and spectral flatness. Each feature was extracted using mirtoolbox functions—*mirenvelope*, *mirrms*, *mirflux*, *mirzerocross*, *mirroughness*, *mirentropy*, *mircentroid*, *mirspread*, *mirrolloff*, *mirflatness*—and adjusted to a sample rate of 125Hz. If the sample number of audio features slightly differed from the EEG data, they were adjusted to match the length of the EEG data: longer samples were cut, and shorter ones were zero-padded. Finally, all audio features were z-scored for normalization.

The SRC calculation was performed using a modified version of a publicly available MATLAB implementation by Dmochowski [1] . SRCs were computed on a per-stimulus basis for each participant. The regularization parameters were set to 7 for both stimuli and EEG data. The representative SRC value for each stimulus and participant was determined by summing the three components with the highest values. As a result, a total of 54 x 10 x 10 (participants x songs x audio features) SRC values were computed.

2.4.3 *Modeling Subjective Experience*

Our analysis used linear mixed-effects models to examine the effects of individual audio features, both in isolation and in conjunction with their corresponding SRC, on subjective music listening experiences: arousal, valence, familiarity, and preference. Separate models were constructed for each dependent variable, with each model incorporating a single audio feature as a fixed effect (AF model). In the case of AFSRC models, compared to AF model, SRC was added as a fixed effect. This approach allowed for a detailed examination of the influence of specific audio features and their neural correlates on listeners' subjective experiences.

[1] https://github.com/dmochow/SRC

The general form of the linear mixed-effects model used in this study is given by:

$$y = X\beta + Z\gamma + \epsilon \tag{1}$$

where y is the vector of observed dependent variables (e.g., arousal, valence), X is the matrix of fixed effects, β represents the coefficients for the fixed effects, Z is the matrix for random effects, γ represents the coefficients for the random effects, and ϵ is the error term.

We fitted two types of models for each dependent variable:

For the audio feature only models, the general form of the model can be represented as:

$$Y_{ij} = \beta_0 + \beta_1 X_{ij} + u_j + \epsilon_{ij} \tag{2}$$

where Y_{ij} is the dependent variable (arousal, valence, familiarity, or preference) for the i-th song listened to by the j-th participant, β_0 is the intercept, β_1 is the fixed effect coefficient of the audio feature X_{ij}, u_j is the random effect for the j-th participant, and ϵ_{ij} is the error term.

For the models with audio feature and SRC as the fixed effects, the equation expands to include the SRC:

$$Y_{ij} = \beta_0 + \beta_1 X_{ij} + \beta_2 S_{ij} + u_j + \epsilon_{ij} \tag{3}$$

where β_2 is the fixed effect coefficient of the SRC S_{ij} related to the audio feature X_{ij}.

The fitting of models was carried out using the `lme4` and `lmerTest` packages in R software. All AF and AF-SRC models were cross-validated using leave-one-subject-out cross-validation. To evaluate the significance of each model, we compared it against a null model predicting the same dependent variable using *anova* function. Specifically, we compared AF models with AFSRC models, again using the *anova* function. When comparing models, it is generally accepted that a difference of 2 or more in AIC values indicates a meaningful difference in model performance [35]. In our experiment results, we also categorized a difference in AIC value of 1.9 as a marginal but meaningful difference. This approach allowed us to quantitatively determine the added value of incorporating EEG-derived SRCs into the predictive models of subjective music listening experience.

3. RESULTS

3.1 Subjective Experience of Fans and non-Fans

Independent samples t-test were conducted to examine the group differences between NCT127 fan group and non-fan group while listening to 10 NCT songs in terms of arousal, valence, familiarity, and preference (Figure 2).

For arousal, there was no significant difference between the fan group (M = 4.54, SD = 1.38) and the non-fan group (M = 4.98, SD = 0.75); $t(25.812) = -1.320, p = .199$.

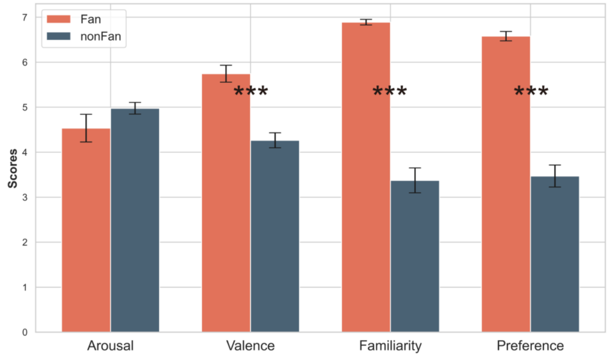

Figure 2: Average subjective ratings by fan and non-fan group. Error bar indicates standard error. *** = p < .001

Valence scores were significantly higher for the fan group (M = 5.75, SD = 0.84) compared to the non-fan group (M = 4.26, SD = 0.97); $t(52) = 5.665$, $p < .001$, indicating that fans experienced more positive emotions towards NCT127 songs.

A significant difference in familiarity with the songs was observed, with the fan group reporting much higher familiarity (M = 6.89, SD = 0.28) than the non-fan group (M = 3.37, SD = 1.61); $t(36.199) = 12.425$, $p < .001$.

Preference ratings were also significantly higher in the fan group (M = 6.58, SD = 0.47) compared to the non-fan group (M = 3.47, SD = 1.42); $t(43.705) = 11.734$, $p < .001$. This result suggests a strong preference for NCT127 music among fans.

Overall, the results indicate that while fans and non-fans do not differ significantly in arousal when listening to NCT127 songs, fans report significantly more positive valence, greater familiarity, and a stronger preference for the NCT127 songs compared to non-fans.

3.2 Predicting Subjective Music Listening Experience with Stimulus-response Correlation

Integrating SRC into the audio feature only models yielded variable results depending on the subjective ratings and audio features. Most importantly, for familiarity, SRC significantly enhanced predictive power of the models across various audio features (Figure 3C). Among ten audio features, SRC correlated with eight audio features showed significant improvement in predicting familiarity.

The prediction of arousal was enhanced from SRC calculated with specific audio feature—spectral flux, spectral centroid, spectral rolloff, and spectral flatness—while roughness was found to negatively impact model performance (Figure 3A). For valence, SRC correlated with spectral flux improved the model performance, whereas sound envelope, RMS, and zero-crossing rate increased the AIC values by 1.9 or more, suggesting reduction in model performance (Figure 3B). In models predicting preference, the addition of SRC related to sound envelope, roughness, spectral centroid, and spectral rolloff resulted in an increase of 1.9 or more in the AIC values, indicating a decline in performance (Figure 3D).

In our analysis of the significance of AF models by comparison to null models, we observed distinct patterns across subjective music listening experiences (Table 1). Specifically, roughness and spectral flatness were key predictors for arousal, while sound envelope, RMS, spectral flux, and zero-crossing rate significantly predicted valence. Familiarity was well predicted by RMS, spectral rolloff, and spectral flatness, and preference was effectively predicted by sound envelope, RMS, spectral flux, zero-crossing rate, spectral entropy, and spectral flatness. Notably, the inclusion of SRC based on RMS, spectral flux, zero-crossing rate, spectral entropy, and spectral flatness did not significantly enhance the performance of models predicting preference (Figure 3D), yet these models demonstrated a good fit using only audio features. For detailed comparisons and summaries of all model fits and cross-validation results, refer to the supplementary materials [2] .

4. DISCUSSION

We compared subjective music listening experiences, specifically focusing on arousal, valence, familiarity, and preference when fans and non-fans of NCT127 listened to the same NCT127 songs. The results showed that valence, familiarity, and preference were significantly higher in the fan group, while there was no significant difference in arousal. Then, we investigated the combined effects of audio features and SRC derived from EEG data on predicting subjective music listening experiences. Through comparing linear mixed-effects models based solely on audio features with those incorporating both audio features and SRC, we revealed that integrating SRC with audio features significantly enhances the predictive power for familiarity. However, the influence of SRC on predictions of arousal and valence showed variation depending on the correlated audio features. The inclusion of few SRC decreased the predictive power of preference.

The notably higher familiarity and preference ratings observed in the NCT127 fan group were anticipated outcomes, aligning with the criteria we set for participant recruitment: participants in the fan group were required to regularly listen to NCT127's music, confirm their attendance at an NCT127 event, and own NCT127-related merchandise.

The absence of significant difference in arousal between groups suggests that arousal ratings were predominantly influenced by the acoustic characteristics of the music, such as tempo and timbre, rather than personal traits [36]. This finding aligns with previous research indicating minimal variability among individuals in arousal ratings for the same musical piece. [37, 38].

Incorporating SRC alongside audio features enhances the predictive accuracy for familiarity. SRC, derived from a hybrid encoding-decoding technique, captures distributed representations in neural response [26]. Since

[2] https://blues95.github.io/ISMIR2024/

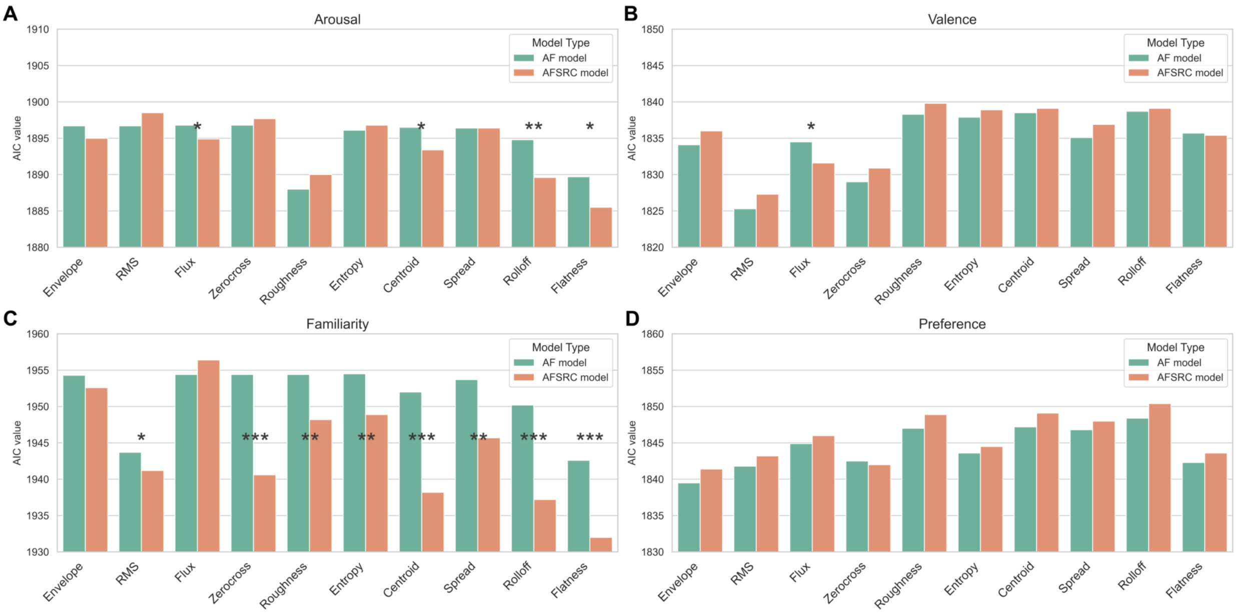

Figure 3: AIC values for each model. (A) Arousal (B) Valence (C) Familiarity (D) Preference. Asterisk symbols indicate the significant improvement of the AFSRC model compared to AF model. Note that the scale of y-axis are different. * = p < .05, ** = p < .01, *** = p < .001.

Table 1: Significance of AF models compared to null models. * = p < .05, ** = p < .01, *** = p < .001.

Audio Feature	Arousal		Valence		Familiarity		Preference	
	AIC	$Pr(>\chi^2)$	AIC	$Pr(>\chi^2)$	AIC	$Pr(>\chi^2)$	AIC	$Pr(>\chi^2)$
Envelope	1896.7	0.682	1834.1	**0.029***	1954.3	0.628	1839.5	**0.002****
RMS	1896.7	0.732	1825.3	**<0.001*****	1943.7	**0.001****	1841.8	**0.007****
Flux	1896.8	0.841	1834.5	**0.037***	1954.4	0.698	1844.9	**0.041***
Zerocross	1896.8	0.925	1829.0	**0.002****	1954.4	0.686	1842.5	**0.010***
Roughness	1888.0	**0.003****	1838.3	0.448	1954.4	0.734	1847.0	0.150
Entropy	1896.1	0.394	1837.9	0.327	1954.5	0.795	1843.6	**0.019***
Centroid	1896.5	0.597	1838.5	0.535	1952.0	0.108	1847.2	0.170
Spread	1896.4	0.531	1835.1	0.052	1953.7	0.357	1846.8	0.127
Rolloff	1894.8	0.150	1838.7	0.751	1950.2	**0.036***	1848.4	0.417
Flatness	1889.7	**0.008****	1835.7	0.076	1942.6	**<0.001*****	1842.3	**0.009****

SRCs in this study were computed by correlating particular audio features with EEG responses, it is possible that audio features of highly familiar music were more effectively represented in neural responses. Familiar music is known to enhance brain activity related to recurring musical patterns and structures [39]. Familiarity may foster better recall of the song, leading to enhanced representation in the brain [40]. Thus, exposure to or familiarity with stimuli may facilitate the processing of specific stimulus features.

In a previous study examining the relationship between audio features and neural responses, Weineck et al. used temporal response function and reliable component analysis to calculate neural synchronization, employing methods distinct from our study [30]. They investigated how synchronization varied with music familiarity, enjoyment, and beat easiness. Their findings indicated that the intensity of neural responses increased with familiar music. While a direct comparison with our study is challenging due to the methodological differences, both studies demonstrate that music familiarity is reflected in the relationship between stimulus (audio features) and response (EEG).

The impact of SRC on the predictions of arousal and valence varied depending on the correlated audio features. In the case of preference, the inclusion of few SRC decreased the model performance, suggesting that emotions or preferences evoked by music may be relatively less dependent on how the audio features are represented in the brain compared to familiarity. Contrary to our findings regarding preference, Pandey et al. demonstrated that stronger SRCs predict increased levels of enjoyment of music [41]. This difference may be due to the selection of features for SRC calculation. Our study used various audio features separately, whereas they used the principal component of 18 audio features for SRC calculation.

The fitting of AF models demonstrated that specific audio features alone can predict subjective music listening experiences. This aligns with the effectiveness of using audio features for training deep learning models in prior MER research.

There are few limitations of this work. First, the demographic composition of our participants, particularly regarding gender distribution, may limit the generalizability of our findings. The process of recruiting fans of a specific artist resulted in a gender imbalance among our participants. Future research should aim to recruit a more balanced participants to enhance the reliability of the results. Second, our analysis only used linear mixed-effects models, making it challenging to generalize the significance of specific audio features in relation to subjective music listening experiences. Since the impact of audio features and SRC on subjective experience may have an inherent nonlinear characteristics, future studies should validate the efficacy of SRC as a learning feature or predictor using a broader range of models, including deep learning-based models capable of capturing nonlinearity. Finally, we only considered ten low-level signal components as audio features in our study. However, the correlations of higher-level audio features, such as chromagrams and various rhythmic features, and EEG might contain unique information about the subjective music listening experience. Therefore, future research should investigate the use of a broader range of audio features, including higher-level audio features.

5. CONCLUSION

This paper explores individual differences in music listening experiences among both fan and non-fan groups of the K-pop idol group NCT127. We aim to demonstrate how responses to the same NCT127 music vary in arousal, valence, familiarity, and preference across different individuals. Furthermore, we investigate the predictive capability of EEG responses, particularly through SRC, regarding subjective music listening experiences. By comparing linear mixed-effects models that solely rely on audio features with those incorporating SRC, our findings underscore the significant role of EEG data in improving the prediction accuracy of music familiarity. This result suggests that using SRC could enable the prediction of individual music listening experiences, which would be challenging using audio features alone.

6. ACKNOWLEDGMENTS

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2023R1A2C100475512).

7. ETHICS STATEMENT

The ethics of the study were approved by the Institutional Review Board of the Korea Advanced Institute of Science and Technology.

8. REFERENCES

[1] R. J. Zatorre, J. L. Chen, and V. B. Penhune, "When the brain plays music: auditory–motor interactions in music perception and production," *Nature Reviews Neuroscience*, vol. 8, no. 7, pp. 547–558, Jul. 2007.

[2] S. Koelsch, "Towards a neural basis of music-evoked emotions," *Trends in Cognitive Sciences*, vol. 14, no. 3, pp. 131–137, 2010. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1364661310000033

[3] ——, "Brain correlates of music-evoked emotions," *Nature Reviews Neuroscience*, vol. 15, no. 3, pp. 170–180, Mar. 2014.

[4] P. Vuust, O. A. Heggli, K. J. Friston, and M. L. Kringelbach, "Music in the brain," *Nature Reviews Neuroscience*, vol. 23, no. 5, pp. 287–305, May 2022.

[5] 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*. ISMIR, Sep. 2015, pp. 538–544. [Online]. Available: https://doi.org/10.5281/zenodo.1416082

[6] E. B. Abrams, E. M. 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*. ISMIR, Nov. 2022, pp. 160–168. [Online]. Available: https://doi.org/10.5281/zenodo.7343078

[7] 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*. ISMIR, Nov. 2020, pp. 566–573. [Online]. Available: https://doi.org/10.5281/zenodo.4245496

[8] N. Rajagopalan and B. Kaneshiro, "Correlation of EEG Responses Reflects Structural Similarity of Choruses in Popular Music," in *Proceedings of the 24th International Society for Music Information Retrieval Conference*. ISMIR, Dec. 2023, pp. 264–271. [Online]. Available: https://doi.org/10.5281/zenodo.10265273

[9] 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, Nov. 2018, pp. 392–399. [Online]. Available: https://doi.org/10.5281/zenodo.1492433

[10] E. Przysinda, T. Zeng, K. Maves, C. Arkin, and P. Loui, "Jazz musicians reveal role of expectancy in human creativity," *Brain and Cognition*, vol. 119, pp. 45–53, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0278262617300994

[11] E. H. Margulis, P. C. M. Wong, C. Turnbull, B. M. Kubit, and J. D. McAuley, "Narratives imagined in response to instrumental music reveal culture-bounded intersubjectivity," *Proceedings of the National Academy of Sciences*, vol. 119, no. 4, p. e2110406119, 2022. [Online]. Available: https://www.pnas.org/doi/abs/10.1073/pnas.2110406119

[12] K. B. Doelling and D. Poeppel, "Cortical entrainment to music and its modulation by expertise," *Proceedings of the National Academy of Sciences*, vol. 112, no. 45, pp. E6233–E6242, 2015. [Online]. Available: https://www.pnas.org/doi/abs/10.1073/pnas.1508431112

[13] Y.-H. Yang and X. Hu, "Cross-cultural Music Mood Classification: A Comparison on English and Chinese Songs." in *Proceedings of the 13th International Society for Music Information Retrieval Conference*. ISMIR, Sep. 2012, pp. 19–24. [Online]. Available: https://doi.org/10.5281/zenodo.1416666

[14] L. Zhang, X. Yang, Y. Zhang, and J. Luo, "Dual Attention-Based Multi-Scale Feature Fusion Approach for Dynamic Music Emotion Recognition," in *Proceedings of the 24th International Society for Music Information Retrieval Conference*. ISMIR, Dec. 2023, pp. 207–214. [Online]. Available: https://doi.org/10.5281/zenodo.10265259

[15] S. Chaki, P. Doshi, S. Bhattacharya, and P. P. Patnaik, "Explaining perceived emotion predictions in music: An attentive approach," in *Proceedings of the 21st International Society for Music Information Retrieval Conference*. ISMIR, Nov. 2020, pp. 150–156. [Online]. Available: https://doi.org/10.5281/zenodo.4245388

[16] Y. E. Kim, E. M. Schmidt, R. Migneco, B. G. Morton, P. Richardson, J. J. Scott, J. A. Speck, and D. Turnbull, "State of the Art Report: Music Emotion Recognition: A State of the Art Review." in *Proceedings of the 11th International Society for Music Information Retrieval Conference*. ISMIR, Sep. 2010, pp. 255–266. [Online]. Available: https://doi.org/10.5281/zenodo.1417945

[17] T. Eerola and J. K. Vuoskoski, "A Review of Music and Emotion Studies: Approaches, Emotion Models, and Stimuli," *Music Perception*, vol. 30, no. 3, pp. 307–340, 02 2013. [Online]. Available: https://doi.org/10.1525/mp.2012.30.3.307

[18] E. B. Lange and K. Frieler, "Challenges and Opportunities of Predicting Musical Emotions with Perceptual and Automatized Features," *Music Perception*, vol. 36, no. 2, pp. 217–242, 12 2018. [Online]. Available: https://doi.org/10.1525/mp.2018.36.2.217

[19] H. Lee, F. Höger, M. Schönwiesner, M. Park, and N. Jacoby, "Cross-cultural Mood Perception in Pop Songs and its Alignment with Mood Detection Algorithms," in *Proceedings of the 22nd International Society for Music Information Retrieval Conference*. ISMIR, Oct. 2021, pp. 366–373. [Online]. Available: https://doi.org/10.5281/zenodo.5625680

[20] J. H. McDermott, A. F. Schultz, E. A. Undurraga, and R. A. Godoy, "Indifference to dissonance in native amazonians reveals cultural variation in music perception," *Nature*, vol. 535, no. 7613, pp. 547–550, Jul. 2016.

[21] E. Mas-Herrero, J. Marco-Pallares, U. Lorenzo-Seva, R. J. Zatorre, and A. Rodriguez-Fornells, "Individual Differences in Music Reward Experiences," *Music Perception*, vol. 31, no. 2, pp. 118–138, 12 2013. [Online]. Available: https://doi.org/10.1525/mp.2013.31.2.118

[22] S. Liljeström, P. N. Juslin, and D. Västfjäll, "Experimental evidence of the roles of music choice, social context, and listener personality in emotional reactions to music," *Psychology of Music*, vol. 41, no. 5, pp. 579–599, 2013. [Online]. Available: https://doi.org/10.1177/0305735612440615

[23] M. Park, K. Hennig-Fast, Y. Bao, P. Carl, E. Pöppel, L. Welker, M. Reiser, T. Meindl, and E. Gutyrchik, "Personality traits modulate neural responses to emotions expressed in music," *Brain Research*, vol. 1523, pp. 68–76, 2013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0006899313007816

[24] K. Mori and M. Iwanaga, "General reward sensitivity predicts intensity of music-evoked chills," *Music Percept.*, vol. 32, no. 5, pp. 484–492, Jun. 2015.

[25] I. Daly, D. Williams, J. Hallowell, F. Hwang, A. Kirke, A. Malik, J. Weaver, E. Miranda, and S. J. Nasuto, "Music-induced emotions can be predicted from a combination of brain activity and acoustic features," *Brain and Cognition*, vol. 101, pp. 1–11, 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0278262615300142

[26] J. P. Dmochowski, J. J. Ki, P. DeGuzman, P. Sajda, and L. C. Parra, "Extracting multidimensional stimulus-response correlations using hybrid encoding-decoding of neural activity," *NeuroImage*, vol. 180, pp. 134–146, 2018, new advances in encoding and decoding of brain signals. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1053811917304299

[27] 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. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S105381192030046X

[28] J. J. Ki, J. P. Dmochowski, J. Touryan, and L. C. Parra, "Neural responses to natural visual motion are spatially selective across the visual field, with selectivity differing across brain areas and task," *European Journal of Neuroscience*, vol. 54, no. 10, pp. 7609–7625, 2021. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/ejn.15503

[29] I. Iotzov and L. C. Parra, "Eeg can predict speech intelligibility," *Journal of Neural Engineering*, vol. 16, no. 3, p. 036008, mar 2019. [Online]. Available: https://dx.doi.org/10.1088/1741-2552/ab07fe

[30] K. Weineck, O. X. Wen, and M. J. Henry, "Neural synchronization is strongest to the spectral flux of slow music and depends on familiarity and beat salience," *eLife*, vol. 11, p. e75515, sep 2022. [Online]. Available: https://doi.org/10.7554/eLife.75515

[31] A. Delorme and S. 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. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0165027003003479

[32] S.-H. Hsu, L. Pion-Tonachini, J. Palmer, M. Miyakoshi, S. Makeig, and T.-P. Jung, "Modeling brain dynamic state changes with adaptive mixture independent component analysis," *NeuroImage*, vol. 183, pp. 47–61, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1053811918306888

[33] L. Pion-Tonachini, K. Kreutz-Delgado, and S. Makeig, "Iclabel: An automated electroencephalographic independent component classifier, dataset, and website," *NeuroImage*, vol. 198, pp. 181–197, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1053811919304185

[34] O. Lartillot and P. Toiviainen, "MIR in Matlab (II): A Toolbox for Musical Feature Extraction from Audio." in *Proceedings of the 8th International Conference on Music Information Retrieval*. ISMIR, Sep. 2018, pp. 127–130. [Online]. Available: https://doi.org/10.5281/zenodo.1417145

[35] S. Müller, J. L. Scealy, and A. H. Welsh, "Model selection in linear mixed models," *Statistical Science*, vol. 28, no. 2, pp. 135–167, 2013. [Online]. Available: http://www.jstor.org/stable/43288485

[36] R. T. Dean, F. Bailes, and E. Schubert, "Acoustic intensity causes perceived changes in arousal levels in music: An experimental investigation," *PLOS ONE*, vol. 6, no. 4, pp. 1–8, 04 2011. [Online]. Available: https://doi.org/10.1371/journal.pone.0018591

[37] S. Yang, C. N. Reed, E. Chew, and M. Barthet, "Examining emotion perception agreement in live music performance," *IEEE Transactions on Affective Computing*, vol. 14, no. 2, pp. 1442–1460, 2023.

[38] B. Gingras, M. M. Marin, E. Puig-Waldmüller, and W. T. Fitch, "The eye is listening: Music-induced arousal and individual differences predict pupillary responses," *Frontiers in Human Neuroscience*, vol. 9, 2015. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fnhum.2015.00619

[39] C. Freitas, E. Manzato, A. Burini, M. J. Taylor, J. P. Lerch, and E. Anagnostou, "Neural correlates of familiarity in music listening: A systematic review and a neuroimaging meta-analysis," *Frontiers in Neuroscience*, vol. 12, 2018. [Online]. Available: https://www.frontiersin.org/journals/neuroscience/articles/10.3389/fnins.2018.00686

[40] V. Vuong, P. Hewan, M. Perron, M. H. Thaut, and C. Alain, "The neural bases of familiar music listening in healthy individuals: An activation likelihood estimation meta-analysis," *Neuroscience Biobehavioral Reviews*, vol. 154, p. 105423, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0149763423003925

[41] P. Pandey, P. S. Bedmutha, K. P. Miyapuram, and D. Lomas, "Stronger correlation of music features with brain signals predicts increased levels of enjoyment," in *2023 IEEE Applied Sensing Conference (APSCON)*, 2023, pp. 1–3.