Decoding Music-Evoked Valence and Arousal: Unraveling the Neural Correlates of Naturalistic Music Characteristics through fMRI

Alexandre Sayal [1,2,3] ✉, **Ana Gabriela Guedes**[1], **Inês Almeida** [1,3], **Daniela Jardim Pereira** [1,4], **César Lima** [5], **Renato Panda** [6,7], **Rui Pedro Paiva** [6], **Teresa Sousa** [1,3,8], **Miguel Castelo-Branco** [1,3,8,9], **Inês Bernardino** [1,3], **Bruno Direito** [1,3,6]

[1]Coimbra Institute for Biomedical Imaging and Translational Research (CIBIT), University of Coimbra; [2]Siemens Healthineers Portugal; [3]Intelligent Systems Associate Laboratory (LASI); [4]ULSCoimbra; [5]ISCTE, University Institute of Lisbon; [6]Center for Informatics and Systems of the University of Coimbra (CISUC), University of Coimbra; [7]Ci2 Smart Cities Research Center, Polytechnic Institute of Tomar; [8]Institute for Nuclear Sciences Applied to Health (ICNAS), University of Coimbra; [9]Faculty of Medicine, University of Coimbra

✉ **For correspondence:** alexandresayal@gmail.com

Data availability: All scripts used for the analyses mentioned in this work can be found here. The anonymized and defaced dataset can be found in BIDS format here. The data management plan regarding this project can be found at Zenodo.

Funding: This research was funded by FCT exploratory project Brainplayback (EXPL/PSI-GER/0948/2021) and CIBIT (UIDB/04950/2020). AS is funded by Siemens Healthineers Portugal and the FCT PhD fellowship (2023.04365.BD).

Competing interests: The authors declare no competing interests.

Abstract

Music can convey basic emotions, such as joy and sadness, and more complex ones, such as tenderness or nostalgia. Its effects on emotion regulation and reward have attracted much attention in cognitive and affective neuroscience. Understanding the underlying neural mechanisms of music-evoked emotions could guide the development of novel technological and individually-tuned neurorehabilitation music-based therapies. This study aims to unravel the relationship between the classification of music excerpts regarding perceived affective states and their associated neural correlates, as measured by fMRI. We used valence and arousal to classify both the stimuli and the affective states perceived by the participants. We acquired fMRI data from 20 participants while listening to 96 musical excerpts a priori classified into four quadrants, considering the valence-arousal model. We first characterized the neural correlates resulting from a GLM analysis of the quadrants defined by valence (positive, negative) and arousal (high, low). Our results highlight the role of neocortical regions, most noticeably the music-specific sub-regions of the auditory cortex and thalamus, as well as regions from the reward network such as the amygdala. Using multivoxel activity patterns corresponding to the four quadrants representation of core affect, we were able to create a computational model that decodes the

quadrant corresponding to the music with significant accuracy, well above a stringent chance level. We further analyzed a set of musical features using regression analysis and explored how they relate to brain activity in valence-, arousal-, reward-, and auditory-related ROIs. The results emphasize the role of expressive features in emotion-related networks. These results contribute to the definition of a relation between music and the neural substrate of music listening and emotions, which is key to developing novel music-based neurorehabilitation strategies.

1 Introduction

Music can elicit a wide range of affective phenomena in the listener, from broader positive or arousing feelings to more specific emotions such as joy, sadness, or nostalgia. The precise mechanisms through which music induces emotions, the characteristics of these emotions, and their connection to other affective processes are ongoing research topics [1]. Understanding this relationship between musical expressivity and the affective response during music listening is essential to develop viable and personalized strategies for emotion regulation [2], [3]. The effects of music on emotion regulation and reward have attracted plenty of attention among cognitive neuroscientists. Functional neuroimaging studies on music and emotion have demonstrated the involvement of emotion-related brain structures, such as the amygdala, nucleus accumbens (NAcc), hypothalamus, hippocampus, insula, cingulate cortex, and orbitofrontal cortex. In particular, [4] meta-analysis showed that music-evoked emotions with positive valence engaged the entire reward network, consisting of the ventral and dorsal striatum (head of the caudate nucleus), amygdala, anterior cingulate cortex (ACC), orbitofrontal cortex (OFC), insula, mediodorsal thalamus (MD), and secondary somatosensory cortex (SII). The involvement of the anterior hippocampus, specific to music paradigms studies, suggests a role in attachment-related emotions associated with promoting social bonding with music [5]–[7]. [5] extended the functional profile of the auditory cortex and hypothesized an emotion-specific functional hub with connections with a broad range of limbic, paralimbic, and neocortical structures. Listening to music involves tracking sound events over time and was found to consistently engage the reward system [8], [9]. Predictive processes are critical in reward during music listening - both the prediction error (predictions about the music itself, i.e., the next chord) and the reward prediction error (how emotionally rewarding a piece of music is) play a role in the pleasure potential of music [10], [11]. Music proficiently exploits our expectations by manipulating melody, rhythm, and patterns [12]. Music emotion recognition, i.e., the relation between specific musical descriptors and emotions, has been previously studied as a correspondence model based on acoustic features (across musical dimensions such as melody, rhythm, harmony, dynamics, timbre, expressivity, texture, and form) and formal, explicit semantic annotations [13], [14]. To develop technological music-based therapeutic applications, identifying the underlying neural mechanisms of music-evoked emotions and how music can convey meaning to the listener is critical [15], [16]. Different theoretical approaches to emotion and music-evoked emotion have been proposed [17], with direct relevance to emotion generation and regulation concepts [18]. Valence and arousal (core affect [19]) can be taken as consciously accessible emergent

properties of more basic physiological processes, reflecting low dimensional descriptive features of all emotional experiences [20]. These dimensional features of emotion represent an interface to study neural responses to emotional experience. The present study addresses the unknown relation between the classification of music excerpts regarding perceived affective states and the associated neural correlates, as measured by functional magnetic resonance imaging (fMRI). To define affective states, low-level dimensional features of emotions conveyed and perceived by music listening were considered. Using multivariate pattern analysis (MVPA) on fMRI data, we aim to address the relation between music characteristics and core affect, provided by valence and arousal dimensions [19], identifying the most important brain regions that encode this information. Previous studies using fMRI were able to successfully decode music-based emotions and explored different elements of the stimuli, such as the duration of music excerpts, decoding accuracy, and the spatial distribution of important voxels [21]–[23]. Here, we aim to identify the brain networks that respond to music-evoked emotions, specifically to low-level dimensions of emotion (valence and arousal), and to understand how acoustical features of the stimuli impact their activity. Ultimately, we aim to establish the connection between music descriptors, music-derived affective states, and brain activity patterns, using machine learning applied to fMRI. We hypothesize that the neural networks involved in affective processes [24], which include emotion-dependent frontal regions, supplementary motor areas, and subcortical regions such as the amygdala, hippocampus, and striatum, in coordination with the auditory cortex, contribute to the mapping of music features and evoked emotions. Most studies have focused on very restricted music stimuli (few genres and/or music excerpts) and usually use mass-univariate approaches. Here, we propose using a heterogeneous dataset, which includes several musical genres, classified and balanced between different quadrants in the arousal and valence plane. These more naturalistic stimuli, combined with the exploratory framework proposed, will allow us to explore the relation between low-level emotional components and patterns of brain activity. Comprehending the mechanisms involved in music listening, characterizing the network supporting music-evoked emotions in music, and defining an encoding model between emotion and dimensions of the music would support the development of novel neurorehabilitative music-based therapies, particularly in disorders involving emotion dysregulation. Computational approaches to neurorehabilitation, such as brain-computer interfaces (BCIs) for neurofeedback, need a more mechanistic understanding of the underlying processes and usually target unspecific, not individually tuned, patterns and parameters.

2 Methods & Materials

2.1 Participants

Twenty individuals (11 females; age range 21-41 years, M = 31, SD = 7) participated in the experiment. All participants gave written informed consent. The study was conducted following the declaration of Helsinki and approved by the Comissão de Ética e Deontologia da Investigação da Faculdade de Psicologia e Ciências de Educação da Universidade de Coimbra. Exclusion criteria included past diagnoses of a neurological or mood disorder and hearing disorders. The mood states of the participants were characterized using the Profile of Mood States (POMS) questionnaire [25],

which includes 42 words describing sensations that people feel in everyday life. Participants were asked to select an answer based on what best corresponds to the way they have been feeling - a scale from 0 (Not at all) to 4 (Very much). Mood state was analyzed according to six subscales - Tension (Mdn= 7.5, range: 2-19), Depression (Mdn= 1.5, range: 0-15), Hostility (Mdn= 2, range: 2-15), Vigor (Mdn= 13.5, range: 5-22), Fatigue (Mdn= 5.5, range: 0-15), and Confusion (Mdn= 5, range: 1-17). A Total Mood Disturbance score revealed a Mdn= 109.5 with a range from 88 to 162. None of the participants was a professional musician.

2.2 Music dataset and acoustical feature analysis

The stimuli consisted of a public dataset, 4Q audio emotion dataset [14], of 900 30-second audio clips, with annotations following Russell's emotion quadrants [19], labeled as Q1 (positive valence, high arousal), Q2 (negative valence, high arousal), Q3 (negative valence, low arousal), and Q4 (positive valence, low arousal). The quadrants are balanced, with precisely 225 music excerpts for each quadrant. The dataset was built using a semi-automatic selection and classification method (AllMusic emotion tags, created and assigned by music experts, were intersected with Warriner's affective - arousal, valence, and dominance - English adjectives list to map the excerpts into Russell's arousal-valence quadrants; following iterations to eliminate non-informative or duplicated tracks, a final human validation was performed; for additional details see [14]). Importantly, the public dataset also includes a total of 1702 acoustic and musical features for each clip. The authors used available audio toolboxes to generate features and created novel, emotionally-relevant audio features based on previous studies. Using statistical classifiers, [14] could discriminate valence and arousal based on these features and extract relevant information regarding music emotion recognition, namely the weight of specific features and musical concepts to each emotion quadrant.

2.3 Procedure

Before scanning, participants filled out the MRI safety questionnaire that verified the participants' compliance with the MRI environment, and the POMS questionnaire. The fMRI session had a total duration of approximately 75 minutes and was divided into four runs. Participants were asked to perform the listening task with their eyes closed. Each participant listened to 96 musical excerpts (24 per run) randomly selected from the public dataset and balanced across the four quadrants. The excerpts were reduced in duration (11.5 seconds) by selecting the first 11.5 seconds available. Within each run, participants listened to 24 music clips, 6 of each quadrant, grouped into two sets of 3 stimuli (each block lasted 36 seconds - 11.5 seconds per excerpt and 0.5 seconds intervals in between). Between the music blocks, participants listened to 12 seconds of ambient sound (inherent to an MRI scanner), then 12 seconds of white noise, and again 12 seconds of ambient sound (Figure 1). The run structure was based on the randomized presentation of music from the four quadrants (two trials) interleaved with ambient sound and noise. Within each block, all stimuli were presented in pseudo-randomized order so that the second pass of music from a quadrant occurred only after the first presentation of all quadrants. The total duration of each run was 10 minutes.

After the scanning session, each participant performed a rating procedure. As emotions are in-

Figure 1. Diagram of the fMRI paradigm trial. Each trial was repeated twice for each of the four quadrants.

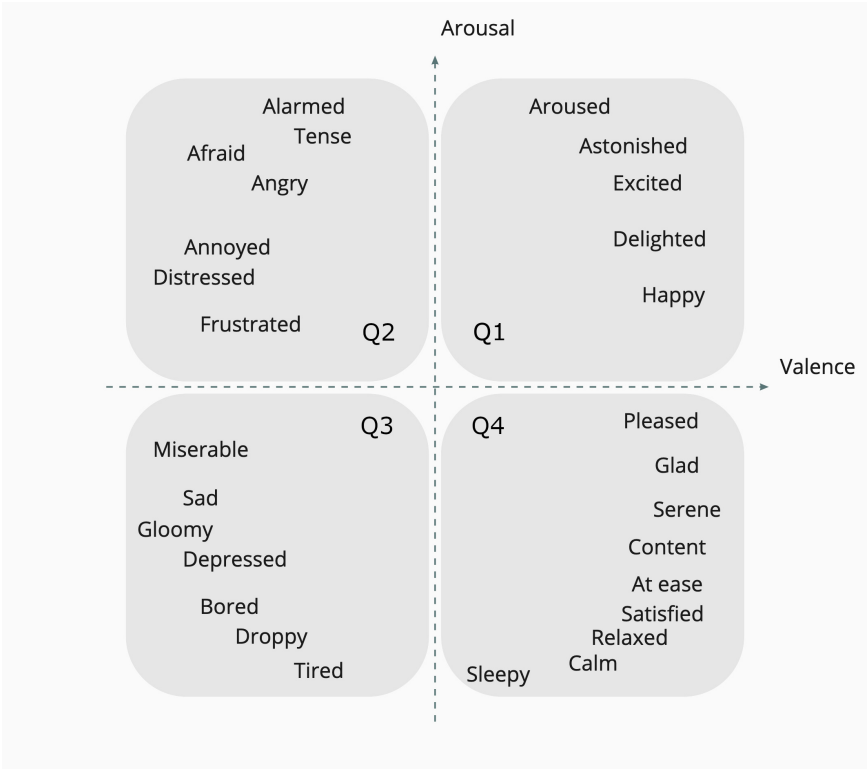

Figure 2. Image of the arousal-valence plane shown to the participants during the rating experiment (the text with the example emotions disappeared after experiment start). Q1 (positive valence, high arousal), Q2 (negative valence, high arousal), Q3 (negative valence, low arousal), and Q4 (positive valence, low arousal).

herently subjective, different people may perceive different emotions when listening to the same song. To account for these differences, while mimicking the previous classification of stimuli according to the Russel's model of affect [19], participants were asked to rate their perceived valence and arousal while listening to the music excerpts that were previously presented during the MRI session (the presentations of the music excerpts was randomized per run). The instruction specifically asked the participants to classify the stimulus in terms of both arousal and valence bipolar dimensions: "Place each of the music excerpts you are going to listen to in the arousal and valence plane considering the emotion that the music transmits to you. The relative position to the center should also be considered". The answer was registered as a mouse click in the 2D arousal-valence plane (Figure 2) and the click should be made when the participant identified the position. A 1-second interval followed each click, with the next music excerpt being then presented. The task was performed with noise-canceling headphones in a quiet room.

2.4 Data acquisition

In the current study, behavior ratings of the music excerpts and neuroimaging data were acquired for each participant, together with a psychological measure of mood. The MRI session and the rating experiment were carried out using Psychopy version 2022.2.5 (Peirce et al. 2019). POMS responses were provided by adapting the 42-item paper questionnaire to a digital form. MR acquisition was performed with a 3 T Siemens Magnetom Prisma fit scanner with a 20-channel head coil at the Institute of Nuclear Sciences Applied to Health, Coimbra. Auditory stimuli were presented using MRI-compatible headphones (Optoacoustics Optoactive II). The scanning session started with the acquisition of two anatomical images: one 3D anatomical magnetization-prepared rapid acquisition gradient echo pulse sequence (repetition time (TR) = 2530 ms, echo time (TE) = 3.5 ms, flip angle (FA) = 7°, 192 slices, voxel size 1.0 × 1.0 × 1.0 mm, field of view (FOV) = 256 × 256 mm) and one T2 space sequence (TR = 3200 ms, TE = 408 ms, 192 slices, voxel size 1.0 × 1.0 × 1.0 mm, FOV = 250 × 250 mm). A total of four functional runs were acquired using a 2D multi-band (MB) gradient-echo echo-planar imaging sequence (TR = 1000 ms, TE = 37 ms, flip angle = 68°, 600 volumes, 66 slices, voxel size 2.0 × 2.0 × 2.0 mm, FOV = 220 mm, MB factor = 6). For correcting image distortions related to magnetic field inhomogeneities, we acquired pairs of spin-echo images with anterior-posterior and posterior-anterior phase encoding polarity with matching geometry and echo-spacing to each of the functional scans (TR = 10250 ms, TE = 73.6 ms). These were acquired before the functional runs. The neuroimaging data were organized according to the BIDS specificatio [26], using BIDSkit (https://github.com/jmtyszka/bidskit) and dcm2niix [27] tools for conversion.

2.5 Data analysis

Ratings of valence and arousal for each of the 24 music excerpts ranged from -1 to 1, given they were presented together in the same plane. We estimated the distribution of the ratings provided by each participant (2D position in the valence-arousal plane) for the music excerpts previously heard during the fMRI session. We report the match between the affective labels and the participants' evaluation. Acquired imaging data were preprocessed using fMRIPrep 20.2.7 [28], which is based on Nipype 1.7.0 [29]. In brief, this pipeline includes slice time correction, head motion correction, unwarping, and normalization to the MNI space. For a step-by-step description of the methods, see supplementary materials. The following sections describe the imaging data analyses performed in Python with Nilearn v0.10.2 [30].

2.5.1 GLM analysis

The first level analysis was based on a single general linear model (GLM) estimated for each participant. The design matrixes included predictors for the music of each quadrant according to the participant's labeling and confound predictors for head motion (six motion parameters + first-order derivatives + powers, a total of 24). Before generating the maps, we applied temporal high pass filtering with a cut-off frequency of 0.007 Hz, spatial smoothing with a Gaussian kernel of full-width at half maximum = 4 mm, and considered a second-order autoregressive model AR(2) as the temporal variance model. The second level analysis consisted of generating the group activation map for the contrast between all the quadrants and noise, corrected for multiple comparisons using

false discovery rate (FDR) (q=0.005). We extract the clusters from this map considering a minimum cluster size of 25 voxels. Based on these clusters, we extracted its z-values for the contrast between each quadrant and noise for each participant. We only considered the voxels inside each cluster mask that were significant at the subject level for the contrast between all quadrants and noise (FDR q=0.05, k>10). We compared the z-values between each pair of quadrants for all clusters using a two-sided Mann-Whitney-Wilcoxon test with FDR correction (q=0.05). All clusters were labeled according to two atlases - AAL3 and Neurosynth - and manually confirmed by a neuroradiologist.

2.5.2 Multivoxel decoding analysis

For each subject, a single GLM was re-performed on the preprocessed functional image, in which each trial of each condition of interest is separated into its own condition within the design matrix. As our conditions of interest are the music excerpts, we defined a predictor for each music excerpt labeled according to the class, i.e. the quadrant in Russell's circumplex, identified by the participant in the rating procedure [31]. The model results in 96 beta maps (24 music excerpts per run, four runs). The individual beta maps were masked using a gray-matter (GM) mask (i.e., non-cortical voxels were ignored) obtained from the fMRIPrep preprocessing of the anatomical images, normalized to zero mean, and scaled to unit variance. Then, a within-subject analysis was performed on these beta maps. Multivariate pattern analysis was carried out using a supervised learning method as the estimator model - a Support Vector Machine (SVM) with a linear kernel and an L1 regularization factor. To solve the multiclass problem, we implemented a one vs. all method. We implemented cross-validation to split the data into different sets, we can then fit the estimator on the training data set and measure an unbiased error on the testing set. To avoid any temporal leakage between training and testing samples (bias due to temporal proximity of samples, particularly relevant in functional data), we implemented a leave-one-run-out (LORO) strategy, where the classifier was trained on 3 of the runs and tested on the remaining one. Leaving out blocks of correlated observations, rather than individual observations, is crucial for non-biased estimates [32].

Prediction accuracy at chance

When performing decoding, the prediction performance of a model can be checked against null distributions or random predictions. For this, we guess a chance level score using simple or random strategies while predicting condition y with X imaging data. In this work, we used dummy predictors implemented in Nilearn, to estimate a chance level score. In summary, the $DummyClassifier$ function makes a prediction that ignores the input features by respecting the class distribution of the training data. This allows us to compare (two-sample t-test) whether the model is better than chance or not while still considering calls imbalance.

Visualizing the decoder's weights

With linear models often used as decoders in neuroimaging, model averaging is appealing as it boils down to averaging weight maps. The weight map spatially represents the model's weights by showing the contribution of each voxel in the image. The weight maps for the SVM model were obtained for each participant, where we highlighted the top 2% of the sorted distribution of the

weights. We used AAL3 to identify the brain structures where clusters were identified [33].

2.5.3 Exploring the importance of acoustic features using regression analysis

Previous research using controlled paradigms played an important role in investigating the neural processing of individual musical features and generating hypotheses for further studies [34]. In this work, we aim to comprehensively explore the relationship between acoustic features (particularly their emotional relevance) and neural correlates. Two factors contribute to this challenge - the number and diversity of the music excerpts used and the set of acoustic features. Here, we considered the top 100 features defined by [14] as the optimal set to discriminate valence and arousal. To appropriately define the relevance of each feature in the valence, arousal, auditory processing, and reward-identified networks, we defined brain masks based on i. large-scale meta-analysis data provided by Neurosynth [35], using the "uniformity-based test" feature (the terms used for the four masks were valence, arousal, primary auditory), and ii. cluster 1 from the meta-analysis presented in [4] as a music-specific reward mask. We intersect these brain masks with the GM mask. To compute the regression coefficients between the beta series and the value of the acoustic features, we used the $DecoderRegressor$ method implemented in Nilearn with ridge regressor as the estimator. It implements a model selection scheme that averages the best models within a cross-validation loop (set as 3 folds). The resulting average model is the one used as a regressor. To assess the statistical significance of the results, we performed a one-sample Wilcoxon test using the coefficients obtained.

3 Results

The following paragraphs present the four sets of findings based on the analysis plan presented above - one regarding the behavior data, and the other three offering different approaches in the analysis of the neuroimaging data. Firstly, we briefly describe the rating task and compare its results with the quadrant labeling of the music dataset. Secondly, we describe the neural correlates of music listening following a GLM approach, and present z-values per condition to characterize valence and arousal in the regions of interest (ROIs). Thirdly, we show the machine learning approach taken to differentiate the four quadrants and present the spatial distribution of the model weights, i.e., the most important voxels for decoding. Finally, we explore the functional patterns of valence, arousal, and primary auditory networks as probed by acoustic features.

3.1 Music excerpt rating in 2D valence and arousal plane

Each participant listened and rated the 24 music excerpts from each quadrant, as identified in the 4Q audio emotion dataset, that were previously played during the fMRI session. Behavioral data results are summarized in the confusion matrix of Figure 3. The results show a degree of mismatch between the labels attributed by the participants and the labels available in the 4Q audio emotion dataset. This is particularly relevant in the low arousal labels, in accordance with the dataset's authors [14]. The participants' labels for Q1 show that 74.8% of the ratings match the 4Q audio emotion dataset. A lower agreement was detected for Q3 music excerpts (low arousal and negative valence), where only 39.0% of the ratings matched with the 4Q audio emotion dataset. The results

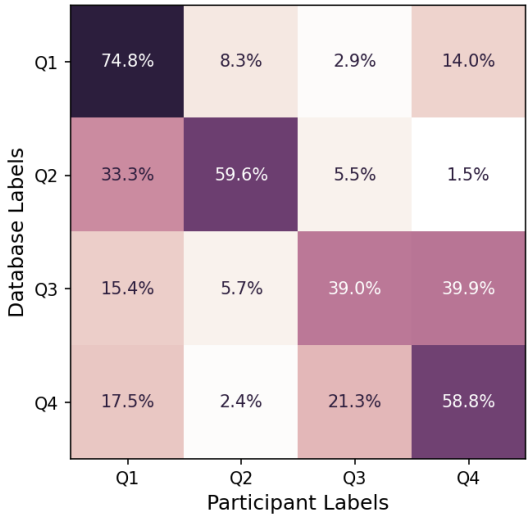

Figure 3. Confusion matrix comparing the labels originally associated with the music clips in the database with the labels posteriorly attributed by the participants to each music clip. The value in each cell indicates the percentage of participant ratings over the total number of music excerpts of each quadrant according to the dataset labels.

indicate participant-specific interpretations of the emotional content of the songs and a mismatch with prior labeling, especially in quadrants Q3 and Q4. Based on the existence of these disparities, which evidence the subjectivity in the emotional interpretation of musical stimuli, we performed all the following analyses using the participants' labels. Figure 4 depicts the overall unbalance between prior labeling and participants' ratings. The results show that Q3 stimuli (considering the 4Q audio emotion dataset labeling) present a neutral perceived valence in our study. On the opposite, the arousal overall pattern follows the prior labeling.

3.2 Neural correlates of music and music-evoked emotions

In Figure 5, we show the group activation map for the contrast between listening to music of all quadrants and noise. We found clusters in the auditory cortex and thalamus, but also regions of the reward network such as the amygdala. The full list of clusters can be found in Table 1 and the z values of each identified cluster divided per quadrant are shown in Figure 6.

3.3 The music's emotional profile identification from its fMRI signature

Due to the re-labeling of the samples (classes defined according to each participant's rating of the music clips), 4 participants were not included in this analysis as there was not enough data from all quadrants to perform the cross-validation method.

3.3.1 Decoding results

The results for the four-class MVPA are shown in Figure 7. The mean overall accuracy was 69.4% (Q1 = 58.1%, Q2 = 75.3%, Q3 = 74.9%, Q4 = 69.3%) and the difference between the overall accuracy and the accuracy of the dummy predictor is statistically significant ($p<0.001$).

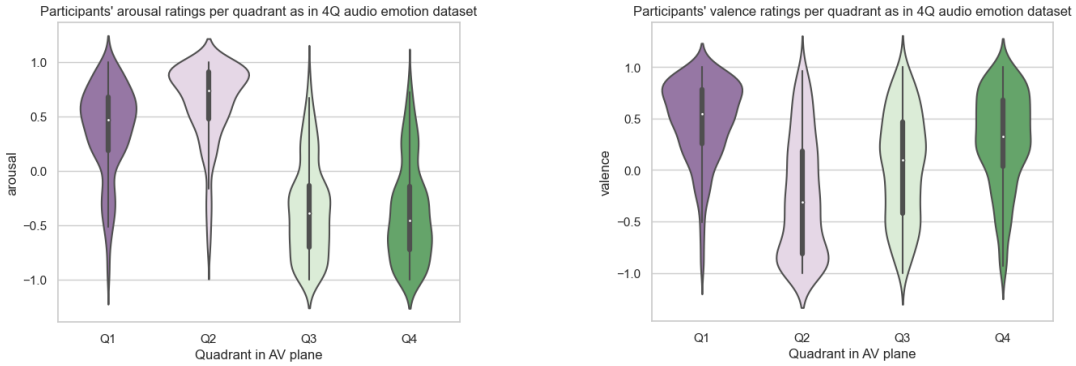

Figure 4. Violin plots displaying the symmetric kernel density estimate and the quartiles of a box plot of the participants' rating for the music of all quadrants in the arousal-valence plane. The values displayed are the Euclidian distance from each axis to the position in the plane where the participants clicked normalized for the furthest report.

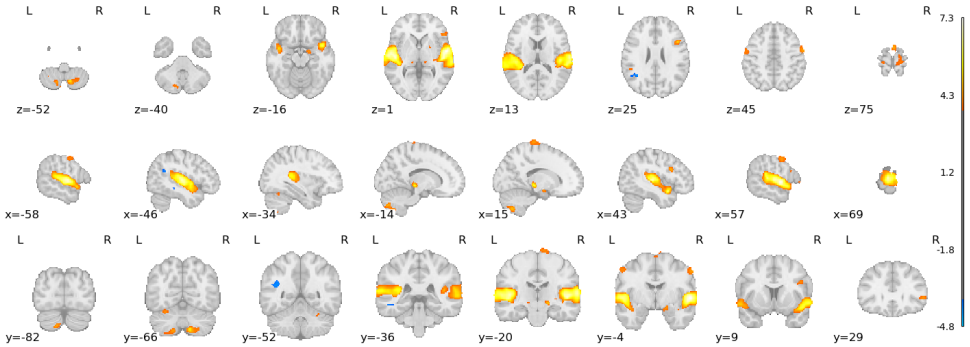

Figure 5. Group activation map for the contrast between all quadrants and noise (z-values FDR corrected, q = 0.005, cluster correction k > 25, N=19).

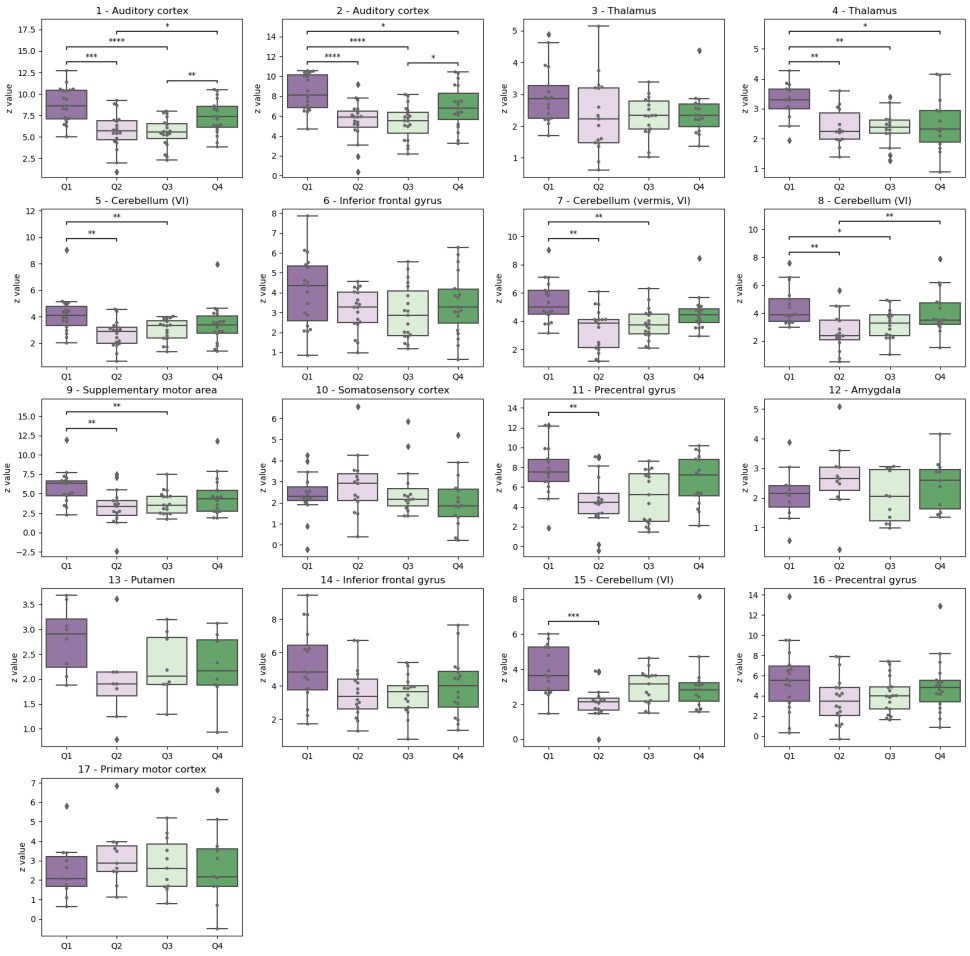

Figure 6. Boxplots for the z values of each identified cluster (Table 1) divided per quadrant. Pairwise statistical comparisons between quadrants are displayed, performed using two-sided Mann-Whitney-Wilcoxon tests with FDR correction (q=0.05).

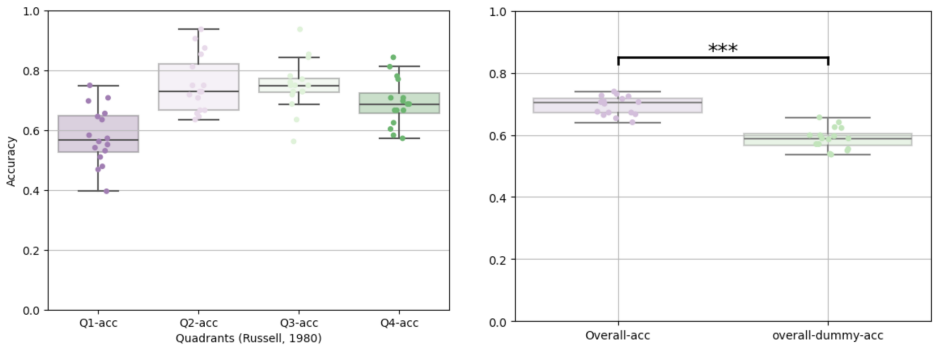

Figure 7. MVPA classification accuracy results. On the left, we display the boxplot of the accuracies achieved per quadrant. On the right, we compare the mean overall accuracy with the accuracy of the dummy predictors (*** - p < 0.001).

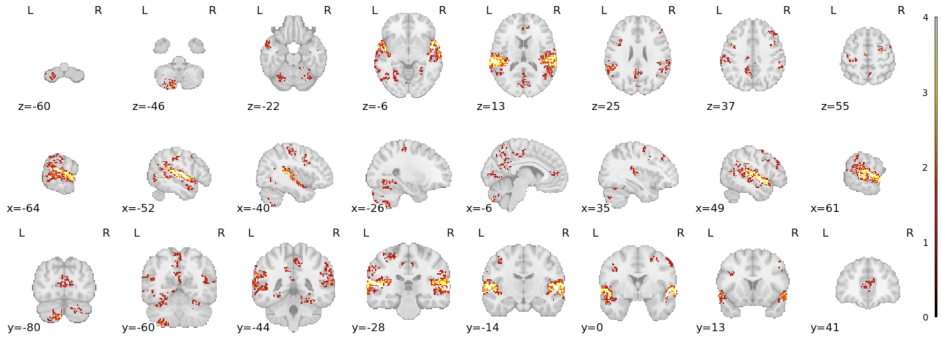

Figure 8. The combination of the top 2% most important voxels for each quadrant to classify valence and arousal categories in the whole-brain analysis (the colorbar represents the number of times that a specific voxel appears in the top 2% - the most important voxels for the classification task).

3.3.2 Visualization of weights

To explore the spatial distribution of the most important voxels to differentiate quadrants, we selected the top 2% most important voxels for each one vs. all model. Figure 8 presents a probability map over the four maps (cluster threshold with 50 voxels). The results show that the most important voxels were predominantly located in the bilateral auditory cortices (Figure 8) for all quadrants. The list of clusters can be found in Table 2. Precentral areas associated with motor functions, the precuneus involved in multiple cognitive tasks and music processing, and the prefrontal areas hub of the working memory network, are also recruited by the models to define a separation hyperplane.

3.4 Acoustic features and brain activation

To explore the impact of different acoustic features in the valence, arousal, and auditory ROIs, we investigated the regression profile of a set of brain masks: (a) valence, (b) arousal, (c) auditory, and (d) reward. To illustrate the differential modulation of these ROIs with all acoustic features, we computed the mean over all voxels of each mask per participant.

3.4.1 Musical features differential engagement in different ROIs

In this analysis, we included a set of 100 candidate acoustic features (based on the previous work by [14]). For a complete overview of these analyses, please consult Supplementary Materials. The exploratory analysis considering the valence mask highlights Expressivity Features, as four of the seven acoustic features presenting statistically significant results are features of expressive techniques in music. In particular, these features are based on Vibrato analysis. Vibrato is an expressive technique used in vocal and instrumental music that consists of a regular pitch oscillation. Its main characteristics are the amount of pitch variation (extent) and the velocity (rate) of this pitch variation. It varies according to different music styles and emotional expressions. Similarly, we see a dominance of Expressivity features when considering the arousal mask. Regarding the auditory mask, we see a significant increase in the number of features associated with brain activation (54 features). Most features are related to Expressivity, Texture, and Tone Color. Considering the

reward mask, we found that Loudness - Spectral and Tonal Dissonance, Pitch (Terhardt) - Chord Change Likelihood, and three features derived from Vibrato were the top-ranked features.

4 Discussion

Our main finding was that valence and arousal can be discriminated above chance level based on the neural representation of music listening. Moreover, this classification is mainly based on bilateral auditory cortices with contributions from the precuneus, frontal, precentral, and cingulate cortices. The GLM analysis showed that the music task activated sensory and emotion-related areas. The analysis of z-values per quadrant showed significant differences in activation of the bilateral auditory cortices, auditory thalamus, and supplementary motor area (SMA), depending on the quadrant. The analysis of acoustic features and its differential recruitment of the valence, arousal, and auditory sensory networks highlighted expressivity features, particularly the Vibrato-related ones as emotion-network modulators. These findings contribute to further comprehending how music relates to emotions and ultimately to developing music-based neurorehabilitative strategies (such as neurofeedback). The inherent ability of music to modulate emotions and regulate mood could indicate that it is a strong candidate for the interface between the participants and their brain activity. Our previous findings support the strategic role of task complexity and feedback valence on the modulation of the network nodes involved in monitoring and feedback control, key variables in neurofeedback framework optimization [36].

4.1 Core affects rating

The distinction between the emotion the listener feels and the one expressed by the music has been the subject of intense debate in the literature. Schubert defined the former as the internal locus of emotion and the latter as the external locus and discussed the relation between the two loci [2]. The wording considered in the rating task is framed in the assessment of the external locus. The short duration of the music excerpts used for the functional task (aimed to explore a wide range of genres and space of acoustical features) was a better fit to study from the point of view of perceived, but not felt, emotions. Nevertheless, [22] have previously concluded that feeling representations emerged within seconds, suggesting the use of shorter musical stimuli. Three main theories are discussed - emotion contagion (both loci are part of the same emotional process, which explains why we tend to feel the emotion that music is conveying) [37], decoding (the listener's decoding will not necessarily be the same and may be susceptible to 'noise'), and dissociation (which explores the complex combination of emotion matches and non-matches, based on mixed cues, between and within loci). Notably, the emotion felt is often rated the same (particularly in the arousal factor) or lower than the corresponding emotion expressed by the stimuli [2]. The selection of a labeled public dataset aimed to appropriately balance the stimuli on the arousal-valence plane. The mismatch identified reinforces the need to tailor all the task elements individually. Several personal and stimulus elements may contribute to variability, such as physiological arousal, personality, and age, as well as musical features. Additionally, the existence of a positive emotional bias, also interpreted as the paradox of tragedy and extensively described in music literature, may also contribute to the shift of the ratings from negative to positive valence quadrants [38]–[40]. All the analyses

presented in this study considered the labels provided by each participant at the cost of stimuli imbalance between classes. This limitation is minimized by the ability of models to adjust weights and decision hyperplanes per class. Low arousal stimuli presented a significant mismatch with the public dataset labels. The ambiguity associated with low arousal has been previously described in music stimuli [41] and interpretation of other emotion-based stimuli, namely facial expressions [42]. Their findings suggest that low-arousal facial expressions are more ambiguous. Moreover, the authors suggest that this is related to a greater activation of an ambiguity/salience system subserved by the amygdala, cerebellum, and dorsal pons. Key variables have been identified as possible confounds in the Music-Emotion relation, such as the wording and detail or task context [43]. The range of wordings found in the literature may benefit from standardization [13].

4.2 Neural correlates of music-induced valence and arousal

4.2.1 GLM and activation per quadrant

The contrast of music vs. noise showed activation clusters in the superior temporal gyrus (STG), auditory thalamus, hippocampus, inferior frontal gyrus, SMA, motor and premotor cortices, amygdala, basal ganglia, inferior frontal gyrus, and the cerebellum. The STG is known as a site of auditory association (and a site of multisensory integration) and thus necessarily plays a crucial role in music listening and music production [44]. As the contrast considers noise as the control condition, at least a component of the activations was not simply due to the sound. In this sense, in accordance with previous findings, the auditory cortex contributes not only to the sensory perception of the auditory stimuli but also to its decoding. The hemodynamic response in the bilateral auditory cortex reflects a differentiated response to valence, as pieces of music labeled from both Q1 and Q4 present stronger responses than the ones labeled with Q2 and Q3. Recent studies have shown that emotional/motivational processes profoundly influence the activity in the auditory cortex and that neurons in the auditory cortex can distinguish between the positive and negative valence of sounds [45]. One of the models proposed for this emotion-informed process in sensory cortices is based on functional connectivity studies that have shown that the reward value of music is correlated with increasing functional connectivity [4]. The auditory thalamus (medial geniculate body) is one of the structures of the auditory pathway and part of the thalamocortical system that provides the anatomical bases for tone and rhythm channels [46]. This structure is also involved in emotional reactions to acoustic stimuli and is part of a thalamocortical circuit that integrates context and content [47]. Frontal activity was found in the inferior frontal gyrus (pars triangularis and frontal operculum), particularly in the right hemisphere. [48] reported on the role of the inferior frontal gyrus in perceiving and feeling emotions during music listening, highlighting an increase in activity during the perceived emotion task. This region was also found in other studies based on MVPA, highlighting its ability to differentiate emotions (fear, happiness, sadness, tenderness, and liking in [49]; fear and joy in [22]). Other activated clusters were found in the somatosensory and motor cortices, premotor regions within the precentral and postcentral gyri, as well as in the SMA. This aligns with previous studies reporting on the engagement of somatomotor regions and the SMA with auditory processing and acoustic cues, emotion perception, and the subjective experience of emotions [44], [50]. Notably, pleasurable music engaging the SMA demonstrates

heightened activation of motor regions, including the SMA and cerebellum, compared to emotionally less evocative music (note that activity during pieces of music labeled as Q1 presents higher values) [51]. Music frequently prompts spontaneous rhythmic movements in listeners. Therefore, the sensorimotor activity observed in the present study may signify the interplay between perception and movement, even without overt physical actions. The GLM also confirmed the contribution of the amygdala in the music-listening task. Moreover, the activation pattern in the amygdala did not show significant differences associated with valence or arousal of the music pieces. Koelsch and colleagues [52] found that brain activity in the amygdala was higher during joy than fear stimuli (after establishing that valence in joy stimuli was significantly higher than for fear). Our finding, on the contrary, accords with the view that the amygdala plays the role of a hub in an emotional coordination structure, engaged in processing both negative and positive stimuli [53]. The reward system plays a central role in the pleasure that most people derive from music. Our results suggest the engagement of this system: the ventral striatum (including the NAcc) integrates the limbic circuit and receives input from the cingulum, OFC, and mesial temporal structures. The dorsal striatum (including the putamen and the caudate nucleus) receives inputs from motor and prefrontal/associative areas and projects to the globus pallidus and substancia nigra. These circuits are highly interlinked and can both lead to dopamine release. As such, the striatum is one of the biological substrates of reward and emotion (pathway of dopaminergic neural responses). Pharmacological manipulation of dopamine causally demonstrated that dopamine mediates musical reward experience, in both positive and negative directions [54]. Several clusters were identified in the cerebellum. The cerebellum has been reported in most decoding studies and these results suggest its relation to music perception and execution [55]. Different parts of the cerebellum have been identified for different aspects of these processing tasks (emotion identification, pleasure in music-listening tasks, etc.) and their individual connections to the cerebral cortex as well as to the basal ganglia. It has been shown, for example, that cerebellar regions are involved in processing specific acoustical features [23], [56]. Altogether, our results reinforce the main structures previously identified in music perception and emotion identification studies. The heterogeneity of our music dataset highlights that this network represents a basis for processing music stimuli, and the quadrant specificity analysis reinforces the role of structures such as the bilateral auditory cortices, auditory thalamus, SMA, and cerebellum in decoding valence and arousal.

4.2.2 Exploring the most important voxels for decoding

The spatial distribution of the most relevant voxels for decoding the different quadrants emphasizes the importance of the auditory cortex, motor areas, cingulate (anterior, middle) gyrus, and frontal middle gyrus for encoding valence and arousal. The results suggest a significant overlap between the main regions identified here and the GLM analysis.

4.3 How does music express different emotions? The neural correlates of acoustical features of music

Music emotion recognition has attracted attention from both music information retrieval and cognitive neuroscience, in a search for a better understanding of the mechanisms associated with

evoked emotions based on music stimuli. Numerous features of music have been reported to be suggestive of discrete emotions [34]. Regarding the reward mask, we found that expressivity, loudness (tonal and spectral dissonance), and pitch were the most associated with the modulation of brain activity. The impact of dissonance and expressive features have been previously found to be relevant in the modulation of reward structures [6], [12], [56]. [57] studied physical voice features and concluded that the amygdala and auditory cortex decode the affective quality of the voice not only by processing the emotional content from previously processed acoustic features but also by processing the acoustic features themselves when these are relevant to the identification of the voice's affective value. Many of the significant features translate aspects of the vibrato. According to an early study, "a good vibrato is a pulsation of pitch, usually accompanied with synchronous pulsation of loudness and timbre, of such extent and rate as to give a pleasing flexibility, tenderness, and richness to the tone" [58]. An irrefutable characteristic of the singing voice [59], it is also extensively used in most musical instruments to address expressivity. Some even argue that vibrato emerged since it mimics the crying voice, and the manipulation of its extent, rate, or intensity may be linked to/identified as different emotions [60]. [21] extracted twenty-one acoustic features capturing timbral, rhythmical, and tonal properties and trained a model to predict brain activity patterns. While achieving 77% accuracy considering voxels within the Heschl's gyrus, secondary auditory regions planum temporale, planum polare, and anterior and posterior STG, evidenced the relation between the modulation of these areas with musical features as roughness - a measure for sensory dissonance, root-mean-square energy - a loudness feature and sub-band flux frequencies between 200 Hz and 1600 Hz and negative loadings of flatness, i.e. a description of the smoothness of the frequency distribution. [61] explored six musical features (labeled by the authors as Fullness, Brightness, Activity, Timbral Complexity, Pulse Clarity, and Key Clarity) and their ability to predict brain responses while listening to music pieces. Areas in the superior temporal gyrus, Heschl's gyrus, Rolandic operculum, and cerebellum contributed to the decoding of timbral features, while for the rhythmic features, the main areas contributing were the bilateral STG, right Heschl's gyrus, and hippocampus. [23] used a spherical searchlight regression analysis to predict brain responses to melody and harmony features. Using searchlight (classification models considering localized spherical masks), statistically significant results were achieved in each of the four cerebral lobes, as well as in the parahippocampal gyrus and the cerebellum. Our results show that expressivity features (particularly the ones based on Vibrato) tend to contribute more to the modulation of valence- and arousal-related brain structures (as defined by a neurosynth meta-analysis). The connectivity between valence structures in the brain mask considered and the auditory cortex (the sensory input) is well-known [47], [62]. In this sense, our data suggests that musical features are directly or indirectly linked to the activation and connectivity of brain areas linked to valence and arousal [63].

4.4 Limitations

We used a public dataset as stimuli to obtain a heterogeneous yet balanced dataset in the arousal-valence plane. This attempt to use valence and arousal ratings from a previous study to establish a balance between quadrants was not successful, as the individual ratings of the participants did not

fully agree, which may have minimized the power of our analysis. Nevertheless, mitigation strategies resulted in sound, comprehensive results. To maximize the heterogeneity and variety of music pieces in the dataset, we selected a relatively short segment of each piece. [22] has previously concluded that feeling representations emerged within seconds. The statistical analysis of the acoustic features was not corrected for multiple comparisons. Considering the number of features analyzed and the exploratory nature of this section of our study, we should limit our interpretation to the trends shown by this analysis. Future confirmatory analyses will require out-of-sample validation or another dataset.

4.5 Conclusion

Our results indicate that several brain regions significantly encode the neural correlates of the perception of valence and arousal in a naturalistic set of music stimuli. Using a heterogeneous dataset, we probed the valence and arousal main brain structures and mapped them. We also explored the potential of several acoustical features to modulate activity in key structures associated with valence, arousal, sensory, and reward networks. The emotion contagion theory (that considers that both loci are part of the same emotional process) reinforces the relation between perceived and felt emotions. However, future works on aiming emotion felt (and neurorehabilitative approaches targeting emotion and mood regulation) should carefully address the optimization of the framework. Previous studies indicate that familiar music elicits stronger emotional reactions and engagement of reward structures [51]. While this is a topic for future research, here we focused on the perception, not the feeling, of the arousal and valence of the music. Future studies, focusing on emotion felt should use longer musical segments and self-selected musical pieces to optimize emotional reaction. Moreover, the arousal-valence plane may not be the optimal emotion framework to explore emotion felt. Aesthetic emotions [64], for example, may represent a more appropriate music emotion model. The connectivity patterns involved in music emotion brain activation should also be addressed in future studies. [9] provided evidence that an increase in the reward value of music is correlated with increasing functional connectivity between the sensory cortex and reward brain structures as the ventral striatum/nucleus accumbens. The understanding of the mechanisms between musical descriptors and specific brain structures (e.g. reward system) is key to optimizing therapeutic strategies. BCIs may take advantage of music stimuli to provide optimal rewarding feedback - neurally informed, immersive, and engaging music feedback depends on the characterization of the link between musical features and brain responses. Several authors have proposed features such as loudness and tempo of music pieces selected a priori as feedback as a regulator of volitional and non-volitional brain activity. Real-time adjustment of specific music features may represent an important alternative for emotion regulation of BCI [65]. Our results suggest that other acoustical features (for example Vibrato) may ultimately represent a better choice to regulate brain activity linked to valence and reward.

4.6 Acknowledgment

We thank all the participants who took part in this study. We also thank the remaining teams members of the Brainplayback project - André Granjo, Carolina Travassos, João Pereira - for the

insightful contributions in the beggining of the project. A last word of gratitude to the team of the MRI Unit of ICNAS - Sónia Afonso, Tânia Lopes - for their support and assistance during the data acquisition.

4.7 Author contributions

Conceived and designed the analysis: AS, RP, RPP, CL, IA, IB, BD; Collected data: AS, BD; Contributed data or analysis tools: RP, RPP, AS, BD, AGG; Performed the analysis: AGG, AS, BD, TBD; Wrote the draft: BD, AS, AGG; Contributed to the manuscript: TBD

References

[1] W. Trost, T. Ethofer, M. Zentner, and P. Vuilleumier, "Mapping Aesthetic Musical Emotions in the Brain," *Cerebral Cortex*, vol. 22, no. 12, pp. 2769–2783, 2012, ISSN: 1460-2199, 1047-3211. DOI: 10.1093/cercor/bhr353. [Online]. Available: https://academic.oup.com/cercor/article-lookup/doi/10.1093/cercor/bhr353.

[2] E. Schubert, "Emotion felt by the listener and expressed by the music: Literature review and theoretical perspectives," *Frontiers in Psychology*, vol. 4, no. DEC, pp. 1–18, 2013, ISSN: 16641078. DOI: 10.3389/fpsyg.2013.00837.

[3] J. Cespedes-Guevara and T. Eerola, "Music communicates affects, not basic emotions - A constructionist account of attribution of emotional meanings to music," *Frontiers in Psychology*, vol. 9, no. FEB, pp. 1–19, 2018, ISSN: 16641078. DOI: 10.3389/fpsyg.2018.00215.

[4] S. Koelsch, "A coordinate-based meta-analysis of music-evoked emotions," *NeuroImage*, vol. 223, p. 117 350, 2020, ISSN: 10538119. DOI: 10.1016/j.neuroimage.2020.117350. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1053811920308363.

[5] S. Koelsch, "Brain correlates of music-evoked emotions," *Nature Reviews Neuroscience*, vol. 15, no. 3, pp. 170–180, 2014, ISSN: 1471003X. DOI: 10.1038/nrn3666.

[6] S. Koelsch, T. Fritz, D. Y. V. Cramon, K. Müller, and A. D. Friederici, "Investigating emotion with music: An fMRI study," *Human Brain Mapping*, vol. 27, no. 3, pp. 239–250, 2006, ISSN: 1065-9471, 1097-0193. DOI: 10.1002/hbm.20180. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1002/hbm.20180.

[7] R. W. Wilkins, D. A. Hodges, P. J. Laurienti, M. Steen, and J. H. Burdette, "Network science and the effects of music preference on functional Brain connectivity: From Beethoven to Eminem," *Scientific Reports*, vol. 4, pp. 1–8, 2014, ISSN: 20452322. DOI: 10.1038/srep06130.

[8] B. P. Gold, M. T. Pearce, E. Mas-Herrero, A. Dagher, and R. J. Zatorre, "Predictability and Uncertainty in the Pleasure of Music: A Reward for Learning?" *The Journal of Neuroscience*, vol. 39, no. 47, pp. 9397–9409, 2019, ISSN: 0270-6474, 1529-2401. DOI: 10.1523/JNEUROSCI.0428-19.2019. [Online]. Available: https://www.jneurosci.org/lookup/doi/10.1523/JNEUROSCI.0428-19.2019.

[9] V. N. Salimpoor, D. H. Zald, R. J. Zatorre, A. Dagher, and A. R. McIntosh, "Predictions and the brain: How musical sounds become rewarding," *Trends in Cognitive Sciences*, vol. 19, no. 2, pp. 86–91, 2015, Publisher: Elsevier Ltd, ISSN: 1879307X. DOI: 10.1016/j.tics.2014.12.001. [Online]. Available: http://dx.doi.org/10.1016/j.tics.2014.12.001.

[10] N. C. Hansen, M. J. Dietz, and P. Vuust, "Commentary: Predictions and the brain: How musical sounds become rewarding," *Frontiers in Human Neuroscience*, vol. 11, no. April, pp. 10–12, 2017, ISSN: 16625161. DOI: 10.3389/fnhum.2017.00168.

[11] D. Huron, *Sweet Anticipation: Music and the Psychology of Expectation*. The MIT Press, 2006, ISBN: 978-0-262-27596-5. DOI: 10.7551/mitpress/6575.001.0001. [Online]. Available: https://direct.mit.edu/books/book/1961/sweet-anticipationmusic-and-the-psychology-of.

[12] H. Chapin, K. Jantzen, J. A. Scott Kelso, F. Steinberg, and E. Large, "Dynamic Emotional and Neural Responses to Music Depend on Performance Expression and Listener Experience," *PLoS ONE*, vol. 5, no. 12, A. Rodriguez-Fornells, Ed., e13812, 2010, ISSN: 1932-6203. DOI: 10.1371/journal.pone.0013812. [Online]. Available: https://dx.plos.org/10.1371/journal.pone.0013812.

[13] J. S. Gomez-Canon, E. Cano, T. Eerola, P. Herrera, X. Hu, Y.-H. Yang, and E. Gomez, "Music Emotion Recognition: Toward new, robust standards in personalized and context-sensitive applications," *IEEE Signal Processing Magazine*, vol. 38, no. 6, pp. 106–114, 2021, ISSN: 1053-5888, 1558-0792. DOI: 10.1109/MSP.2021.3106232. [Online]. Available: https://ieeexplore.ieee.org/document/9591555/.

[14] R. Panda, R. Malheiro, and R. P. Paiva, "Novel Audio Features for Music Emotion Recognition," *IEEE Transactions on Affective Computing*, vol. 11, no. 4, pp. 614–626, 2020, ISSN: 1949-3045, 2371-9850. DOI: 10.1109/TAFFC.2018.2820691. [Online]. Available: https://ieeexplore.ieee.org/document/8327886/.

[15] N. K. Flaig and E. W. Large, "Dynamic musical communication of core affect," *Frontiers in Psychology*, vol. 5, 2014, ISSN: 1664-1078. DOI: 10.3389/fpsyg.2014.00072. [Online]. Available: http://journal.frontiersin.org/article/10.3389/fpsyg.2014.00072/abstract.

[16] P. N. Juslin and D. Västfjäll, "Emotional responses to music: The need to consider underlying mechanisms," *Behavioral and Brain Sciences*, vol. 31, no. 5, pp. 559–575, 2008, ISSN: 0140-525X, 1469-1825. DOI: 10.1017/S0140525X08005293. [Online]. Available: https://www.cambridge.org/core/product/identifier/S0140525X08005293/type/journal_article.

[17] L. A. Warrenburg, "Comparing musical and psychological emotion theories.," *Psychomusicology: Music, Mind, and Brain*, vol. 30, no. 1, pp. 1–19, 2020, ISSN: 2162-1535, 0275-3987. DOI: 10.1037/pmu0000247. [Online]. Available: https://doi.apa.org/doi/10.1037/pmu0000247.

[18] J. J. Gross and L. Feldman Barrett, "Emotion Generation and Emotion Regulation: One or Two Depends on Your Point of View," *Emotion Review*, vol. 3, no. 1, pp. 8–16, 2011, ISSN: 1754-0739, 1754-0747. DOI: 10.1177/1754073910380974. [Online]. Available: http://journals.sagepub.com/doi/10.1177/1754073910380974.

[19] J. A. Russell, "A circumplex model of affect," *Journal of Personality and Social Psychology*, vol. 39, no. 6, pp. 1161–1178, 1980, ISSN: 00223514. DOI: 10.1037/h0077714.

[20] L. F. Barrett, "The theory of constructed emotion: An active inference account of interoception and categorization," *Social cognitive and affective neuroscience*, vol. 12, no. 1, pp. 1–23, 2016, ISSN: 17495024. DOI: 10.1093/scan/nsw154.

[21] S. Hoefle, A. Engel, R. Basilio, V. Alluri, P. Toiviainen, M. Cagy, and J. Moll, "Identifying musical pieces from fMRI data using encoding and decoding models," *Scientific Reports*, vol. 8, no. 1, pp. 1–13, 2018, ISSN: 20452322. DOI: 10.1038/s41598-018-20732-3.

[22] S. Koelsch, V. K. Cheung, S. Jentschke, and J. D. Haynes, "Neocortical substrates of feelings evoked with music in the ACC, insula, and somatosensory cortex," *Scientific Reports*, vol. 11, no. 1, pp. 1–11, 2021, Publisher: Nature Publishing Group UK ISBN: 0123456789, ISSN: 20452322. DOI: 10.1038/s41598-021-89405-y. [Online]. Available: https://doi.org/10.1038/s41598-021-89405-y.

[23] M. A. Casey, "Music of the 7Ts: Predicting and decoding multivoxel fMRI responses with acoustic, schematic, and categorical music features," *Frontiers in Psychology*, vol. 8, no. JUL, pp. 1–11, 2017, ISSN: 16641078. DOI: 10.3389/fpsyg.2017.01179.

[24] L. F. Barrett and E. Bliss-Moreau, "Chapter 4 Affect as a Psychological Primitive," in *Advances in Experimental Social Psychology*, vol. 41, Elsevier, 2009, pp. 167–218, ISBN: 978-0-12-374472-2. DOI: 10.1016/S0065-2601(08)00404-8. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0065260108004048.

[25] M. Faro Viana, P. Almeida, and R. C. Santos, "Adaptação portuguesa da versão reduzida do Perfil de Estados de Humor – POMS," *Análise Psicológica*, vol. 19, no. 1, pp. 77–92, 2001, arXiv: 0195-9131/02/3402-0364/0 ISBN: 1948-7134, ISSN: 1646-6020. DOI: 10.14417/ap.345. [Online]. Available: http://publicacoes.ispa.pt/index.php/ap/article/view/345.

[26] K. J. Gorgolewski, T. Auer, V. D. Calhoun, R. C. Craddock, S. Das, E. P. Duff, G. Flandin, S. S. Ghosh, T. Glatard, Y. O. Halchenko, D. A. Handwerker, M. Hanke, D. Keator, X. Li, Z. Michael, C. Maumet, B. N. Nichols, T. E. Nichols, J. Pellman, J. B. Poline, A. Rokem, G. Schaefer, V. Sochat, W. Triplett, J. A. Turner, G. Varoquaux, and R. A. Poldrack, "The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments," *Scientific Data*, vol. 3, pp. 1–9, 2016, ISBN: 9780387734118, ISSN: 20524463. DOI: 10.1038/sdata.2016.44.

[27] X. Li, P. S. Morgan, J. Ashburner, J. Smith, and C. Rorden, "The first step for neuroimaging data analysis: DICOM to NIfTI conversion," *Journal of Neuroscience Methods*, vol. 264, pp. 47–56, 2016, ISSN: 0165-0270. DOI: 10.1016/j.jneumeth.2016.03.001. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0165027016300073.

[28] O. Esteban, C. J. Markiewicz, R. W. Blair, C. A. Moodie, A. I. Isik, A. Erramuzpe, J. D. Kent, M. Goncalves, E. DuPre, M. Snyder, H. Oya, S. S. Ghosh, J. Wright, J. Durnez, R. A. Poldrack, and K. J. Gorgolewski, "fMRIPrep: A robust preprocessing pipeline for functional MRI," *Nature Methods*, vol. 16, no. 1, pp. 111–116, 2019, ISSN: 1548-7091, 1548-7105. DOI: 10.1038/s41592-018-0235-4. [Online]. Available: https://www.nature.com/articles/s41592-018-0235-4.

[29] K. Gorgolewski, C. D. Burns, C. Madison, D. Clark, Y. O. Halchenko, M. L. Waskom, and S. S. Ghosh, "Nipype: A Flexible, Lightweight and Extensible Neuroimaging Data Processing Framework in Python," *Frontiers in Neuroinformatics*, vol. 5, 2011, ISSN: 1662-5196. DOI: 10.3389/fninf.2011.00013. [Online]. Available: http://journal.frontiersin.org/article/10.3389/fninf.2011.00013/abstract.

[30] A. Abraham, F. Pedregosa, M. Eickenberg, P. Gervais, A. Mueller, J. Kossaifi, A. Gramfort, B. Thirion, and G. Varoquaux, "Machine learning for neuroimaging with scikit-learn," *Frontiers in Neuroinformatics*, vol. 8, 2014, ISSN: 1662-5196. DOI: 10.3389/fninf.2014.00014. [Online]. Available: http://journal.frontiersin.org/article/10.3389/fninf.2014.00014/abstract.

[31] J. A. Mumford, B. O. Turner, F. G. Ashby, and R. A. Poldrack, "Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses," *NeuroImage*, vol. 59, no. 3, pp. 2636–2643, 2012, Publisher: Elsevier Inc., ISSN: 10538119. DOI: 10.1016/j.neuroimage.2011.08.076. [Online]. Available: http://dx.doi.org/10.1016/j.neuroimage.2011.08.076.

[32] G. Varoquaux, P. R. Raamana, D. A. Engemann, A. Hoyos-Idrobo, Y. Schwartz, and B. Thirion, "Assessing and tuning brain decoders: Cross-validation, caveats, and guidelines," *NeuroImage*, vol. 145, pp. 166–179, 2017, ISSN: 10538119. DOI: 10.1016/j.neuroimage.2016.10.038. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S105381191630595X.

[33] E. T. Rolls, C.-C. Huang, C.-P. Lin, J. Feng, and M. Joliot, "Automated anatomical labelling atlas 3," *NeuroImage*, vol. 206, p. 116 189, 2020, ISSN: 10538119. DOI: 10.1016/j.neuroimage.2019.116189. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1053811919307803.

[34] P. N. Juslin and P. Laukka, "Expression, Perception, and Induction of Musical Emotions: A Review and a Questionnaire Study of Everyday Listening," *Journal of New Music Research*, vol. 33, no. 3, pp. 217–238, 2004, ISSN: 0929-8215, 1744-5027. DOI: 10.1080/0929821042000317813. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/0929821042000317813.

[35] T. Yarkoni, R. A. Poldrack, T. E. Nichols, D. C. Van Essen, and T. D. Wager, "Large-scale automated synthesis of human functional neuroimaging data," *Nature Methods*, vol. 8, no. 8, pp. 665–670, 2011, ISBN: 1662-5196, ISSN: 15487091. DOI: 10.1038/nmeth.1635.

[36] B. Direito, M. Ramos, J. Pereira, A. S. Campos, T. Sousa, and M. de Sá e Sousa Castelo-Branco, "Directly exploring the neural correlates of feedback-related reward saliency during fMRI-based Neurofeedback," *Frontiers in Human Neuroscience*, vol. 14, no. February, p. 578 119, 2021, arXiv: 1011.1669v3 ISBN: 9788578110796, ISSN: 1098-6596. DOI: 10.3389/fnhum.2020.578119.

[37] A. Gabrielsson, "Emotion perceived and emotion felt: Same or different?" *Musicae Scientiae*, vol. 5, no. 1_suppl, pp. 123–147, 2001, ISSN: 1029-8649, 2045-4147. DOI: 10.1177/10298649020050S105. [Online]. Available: http://journals.sagepub.com/doi/10.1177/10298649020050S105.

[38] T. Eerola, J. K. Vuoskoski, H.-R. Peltola, V. Putkinen, and K. Schäfer, "An integrative review of the enjoyment of sadness associated with music," *Physics of Life Reviews*, vol. 25, pp. 100–121, 2018, ISSN: 15710645. DOI: 10.1016/j.plrev.2017.11.016. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S157106451730163X.

[39] M. E. Sachs, A. Damasio, and A. Habibi, "The pleasures of sad music: A systematic review," *Frontiers in Human Neuroscience*, vol. 9, 2015, ISSN: 1662-5161. DOI: 10.3389/fnhum.2015.00404. [Online]. Available: http://journal.frontiersin.org/Article/10.3389/fnhum.2015.00404/abstract.

[40] J. S. Blasco-Magraner, G. Bernabé-Valero, P. Marín-Liébana, and A. M. Botella-Nicolás, "Changing positive and negative affects through music experiences: A study with university students," *BMC Psychology*, vol. 11, no. 1, p. 76, 2023, ISSN: 2050-7283. DOI: 10.1186/s40359-023-01110-9. [Online]. Available: https://bmcpsychology.biomedcentral.com/articles/10.1186/s40359-023-01110-9.

[41] Y. Hong, C.-J. Chau, and A. Horner, "An Analysis of Low-Arousal Piano Music Ratings to Uncover What Makes Calm and Sad Music So Difficult to Distinguish in Music Emotion Recognition," *Journal of the Audio Engineering Society*, vol. 65, no. 4, pp. 304–320, 2017, ISSN: 15494950. DOI: 10.17743/jaes.2017.0001. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=18563.

[42] A. J. Gerber, J. Posner, D. Gorman, T. Colibazzi, S. Yu, Z. Wang, A. Kangarlu, H. Zhu, J. Russell, and B. S. Peterson, "An Affective Circumplex Model of Neural Systems Subserving Valence, Arousal, & Cognitive Overlay During the Appraisal of Emotional Faces," 2009.

[43] V. J. Konečni, "Does music induce emotion? A theoretical and methodological analysis.," *Psychology of Aesthetics, Creativity, and the Arts*, vol. 2, no. 2, pp. 115–129, 2008, ISSN: 1931-390X, 1931-3896. DOI: 10.1037/1931-3896.2.2.115. [Online]. Available: http://doi.apa.org/getdoi.cfm?doi=10.1037/1931-3896.2.2.115.

[44] V. Pando-Naude, A. Patyczek, L. Bonetti, and P. Vuust, "An ALE meta-analytic review of top-down and bottom-up processing of music in the brain," *Scientific Reports*, vol. 11, no. 1, pp. 1–15, 2021, Publisher: Nature Publishing Group UK ISBN: 0123456789, ISSN: 20452322. DOI: 10.1038/s41598-021-00139-3. [Online]. Available: https://doi.org/10.1038/s41598-021-00139-3.

[45] G. Concina, A. Renna, A. Grosso, and B. Sacchetti, "The auditory cortex and the emotional valence of sounds," *Neuroscience & Biobehavioral Reviews*, vol. 98, pp. 256–264, 2019, ISSN: 01497634. DOI: 10.1016/j.neubiorev.2019.01.018. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0149763418307619.

[46] G. Musacchia, E. W. Large, and C. E. Schroeder, "Thalamocortical mechanisms for integrating musical tone and rhythm," *Hearing Research*, vol. 308, pp. 50–59, 2014, ISSN: 03785955. DOI: 10.1016/j.heares.2013.09.017. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0378595513002402.

[47] S. Koelsch and S. Skouras, "Functional centrality of amygdala, striatum and hypothalamus in a "small-world" network underlying joy: An fMRI study with music: A Neural Network Underlying Joy," *Human Brain Mapping*, vol. 35, no. 7, pp. 3485–3498, 2014, ISSN: 10659471. DOI: 10.1002/hbm.22416. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1002/hbm.22416.

[48] K.-i. Tabei, "Inferior Frontal Gyrus Activation Underlies the Perception of Emotions, While Precuneus Activation Underlies the Feeling of Emotions during Music Listening," *Behavioural Neurology*, vol. 2015, pp. 1–6, 2015, ISSN: 0953-4180, 1875-8584. DOI: 10.1155/2015/529043. [Online]. Available: http://www.hindawi.com/journals/bn/2015/529043/.

[49] V. Putkinen, S. Nazari-Farsani, K. Seppälä, T. Karjalainen, L. Sun, H. K. Karlsson, M. Hudson, T. T. Heikkilä, J. Hirvonen, and L. Nummenmaa, "Decoding Music-Evoked Emotions in the Auditory and Motor Cortex," *Cerebral Cortex*, vol. 31, no. 5, pp. 2549–2560, 2021, ISSN: 1047-3211, 1460-2199. DOI: 10.1093/cercor/bhaa373. [Online]. Available: https://academic.oup.com/cercor/article/31/5/2549/6046263.

[50] C. F. Lima, S. Krishnan, and S. K. Scott, "Roles of Supplementary Motor Areas in Auditory Processing and Auditory Imagery," *Trends in Neurosciences*, vol. 39, no. 8, pp. 527–542, 2016, Publisher: Elsevier Ltd, ISSN: 1878108X. DOI: 10.1016/j.tins.2016.06.003. [Online]. Available: http://dx.doi.org/10.1016/j.tins.2016.06.003.

[51] A. J. Blood and R. J. Zatorre, "Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion," *Proceedings of the National Academy of Sciences*, vol. 98, no. 20, pp. 11 818–11 823, 2001, ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.191355898. [Online]. Available: https://pnas.org/doi/full/10.1073/pnas.191355898.

[52] S. Koelsch, S. Skouras, T. Fritz, P. Herrera, C. Bonhage, M. B. Küssner, and A. M. Jacobs, "The roles of superficial amygdala and auditory cortex in music-evoked fear and joy," *NeuroImage*, vol. 81, pp. 49–60, 2013, Publisher: Elsevier Inc., ISSN: 10538119. DOI: 10.1016/j.neuroimage.2013.05.008. [Online]. Available: http://dx.doi.org/10.1016/j.neuroimage.2013.05.008.

[53] K. A. Lindquist, A. B. Satpute, T. D. Wager, J. Weber, and L. F. Barrett, "The Brain Basis of Positive and Negative Affect: Evidence from a Meta-Analysis of the Human Neuroimaging Literature," *Cerebral Cortex*, vol. 26, no. 5, pp. 1910–1922, 2016, ISSN: 1047-3211, 1460-2199. DOI: 10.1093/cercor/bhv001. [Online]. Available: https://academic.oup.com/cercor/article-lookup/doi/10.1093/cercor/bhv001.

[54] L. Ferreri, E. Mas-Herrero, R. J. Zatorre, P. Ripollés, A. Gomez-Andres, H. Alicart, G. Olivé, J. Marco-Pallarés, R. M. Antonijoan, M. Valle, J. Riba, and A. Rodriguez-Fornells, "Dopamine modulates the reward experiences elicited by music," *Proceedings of the National Academy of Sciences*, vol. 116, no. 9, pp. 3793–3798, 2019, ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1811878116. [Online]. Available: https://pnas.org/doi/full/10.1073/pnas.1811878116.

[55] 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, 2007, ISSN: 1471-003X, 1471-0048. DOI: 10.1038/nrn2152. [Online]. Available: https://www.nature.com/articles/nrn2152.

[56] 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*, vol. 59, no. 4, pp. 3677–3689, 2012, Publisher: Elsevier Inc., ISSN: 10538119. DOI:

10.1016/j.neuroimage.2011.11.019. [Online]. Available: http://dx.doi.org/10.1016/j.neuroimage.2011.11.019.

[57] A. Pannese, D. Grandjean, and S. Frühholz, "Amygdala and auditory cortex exhibit distinct sensitivity to relevant acoustic features of auditory emotions," *Cortex*, vol. 85, pp. 116–125, 2016, ISSN: 00109452. DOI: 10.1016/j.cortex.2016.10.013. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0010945216303045.

[58] C. E. Seashore, *The Psychology of Music*. 1937. [Online]. Available: https://journals.sagepub.com/doi/10.2307/3385146.

[59] C. Dromey, S. O. Holmes, J. A. Hopkin, and K. Tanner, "The Effects of Emotional Expression on Vibrato," *Journal of Voice*, vol. 29, no. 2, pp. 170–181, 2015, ISSN: 08921997. DOI: 10.1016/j.jvoice.2014.06.007. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S089219971400126X.

[60] F. Eyben, G. L. Salomão, J. Sundberg, K. R. Scherer, and B. W. Schuller, "Emotion in the singing voice—a deeperlook at acoustic features in the light ofautomatic classification," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2015, no. 1, p. 19, 2015, ISSN: 1687-4722. DOI: 10.1186/s13636-015-0057-6. [Online]. Available: https://asmp-eurasipjournals.springeropen.com/articles/10.1186/s13636-015-0057-6.

[61] 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*, vol. 88, pp. 170–180, 2014, Publisher: Elsevier Inc., ISSN: 10538119. DOI: 10.1016/j.neuroimage.2013.11.017. [Online]. Available: http://dx.doi.org/10.1016/j.neuroimage.2013.11.017.

[62] V. Menon and D. J. Levitin, "The rewards of music listening: Response and physiological connectivity of the mesolimbic system," *NeuroImage*, vol. 28, no. 1, pp. 175–184, 2005, ISSN: 10538119. DOI: 10.1016/j.neuroimage.2005.05.053.

[63] S.-G. Kim, K. Mueller, J. Lepsien, T. Mildner, and T. H. Fritz, "Brain networks underlying aesthetic appreciation as modulated by interaction of the spectral and temporal organisations of music," *Scientific Reports*, vol. 9, no. 1, p. 19 446, 2019, ISSN: 2045-2322. DOI: 10.1038/s41598-019-55781-9. [Online]. Available: https://www.nature.com/articles/s41598-019-55781-9.

[64] M. Zentner, D. Grandjean, and K. R. Scherer, "Emotions evoked by the sound of music: Characterization, classification, and measurement.," *Emotion*, vol. 8, no. 4, pp. 494–521, 2008, ISSN: 1931-1516, 1528-3542. DOI: 10.1037/1528-3542.8.4.494. [Online]. Available: http://doi.apa.org/getdoi.cfm?doi=10.1037/1528-3542.8.4.494.

[65] K. Takabatake, N. Kunii, H. Nakatomi, S. Shimada, K. Yanai, M. Takasago, and N. Saito, "Musical Auditory Alpha Wave Neurofeedback: Validation and Cognitive Perspectives," *Applied Psychophysiology and Biofeedback*, vol. 46, no. 4, pp. 323–334, 2021, ISSN: 1090-0586, 1573-3270. DOI: 10.1007/s10484-021-09507-1. [Online]. Available: https://link.springer.com/10.1007/s10484-021-09507-1.

Table 1. Cluster table of the group activation map contrasting music from all quadrants and noise (FDR corrected, q = 0.005, cluster correction k > 25, N=19). The labels for major clusters were obtained from the AAL3 atlas and the Neurosynth association tab.

Cluster ID	X	Y	Z	Peak z value	Cluster Size (mm3)	Labels
1	58	-5	2	7.29	34968	Temporal superior sulcus, auditory cortex
1a	50	-15	4	7.05		
1b	66	-29	8	6.89		
1c	44	-25	6	6.82		
2	-59	-11	4	7.01	34376	Temporal superior sulcus, auditory cortex
2a	-49	-17	2	6.95		
2b	-45	-23	8	6.75		
2c	-51	-1	-3	6.43		
3	16	-25	-5	5.84	952	Thalamus
4	-15	-27	-7	5.50	816	Thalamus
5	12	-67	-53	5.35	2376	Cerebellum (VI)
5a	30	-61	-47	4.89		
5b	12	-77	-47	4.56		
6	48	16	22	4.97	1040	Inferior frontal gyrus
6a	60	16	24	3.81		
7	-13	-81	-45	4.65	2416	Cerebellum (vermis, VI)
7a	-15	-69	-53	4.54		
7b	-29	-61	-57	4.30		
7c	-15	-75	-51	4.20		
8	-29	-61	-25	4.57	984	Cerebellum (VI)
9	2	2	76	4.56	1272	Supplementary motor area
10	14	-25	82	4.50	2104	Somatosensory cortex
11	60	-5	46	4.40	1096	Precentral gyrus (face, mouth)
12	18	-7	-15	4.40	360	Amygdala
13	24	2	6	4.37	312	Putamen
14	52	30	2	4.32	520	Inferior frontal gyrus
15	30	-57	-25	4.23	408	Cerebellum (VI)
16	-57	-5	48	4.18	800	Precentral gyrus (face, mouth)
17	-17	-29	78	3.95	264	Primary motor cortex

Table 2. Description of the clusters identified in the map of Figure 8, displaying the coordinates of the peak voxel, the number of quadrants at that coordinate, and the labels from the AAL3 atlas and Neurosynth associations.

X	Y	Z	#Q at peak	Cluster Size (mm3)	Labels
-65	-3	-3	4	31904	Temporal superior sulcus, auditory cortex
56	-41	12	4	29896	Temporal superior sulcus, auditory cortex
-49	-67	-9	2	1936	Temporal inferior, visual ventral stream
-47	8	38	2	1760	Precentral gyrus, language
-33	-31	56	2	1552	Precentral gyrus, motor
-27	-53	-9	2	3032	Lingual gyrus
-13	-37	44	2	3888	Middle cingulate cortex
-11	-55	40	2	6664	Precuneus
-3	44	8	1	1184	Anterior cingulate cortex
24	-45	-19	1	872	Fusiform gyrus
44	32	34	1	1400	Frontal middle gyrus, pre-frontal cortex, working memory
44	6	40	1	800	Frontal middle gyrus, pre-frontal cortex
56	-1	46	1	816	Precentral gyrus, motor