Decoding Musical Valence and Arousal: Exploring the Neural Correlates of Music-Evoked Emotions and the Role of Expressivity Features

Alexandre Sayal*, Ana Gabriela Guedes, Inês Almeida, Daniela Jardim Pereira, César F. Lima, Renato Panda, Rui Pedro Paiva, Teresa Sousa, Miguel Castelo-Branco, Inês Bernardino, Bruno Direito

Abstract—Music conveys both basic emotions, like joy and sadness, and complex ones, such as tenderness and nostalgia. Its effects on emotion regulation and reward have attracted much research attention, as the neural correlates of music-evoked emotions may inform neurorehabilitation interventions. Here, we used fMRI to decode and examine the neural correlates of perceived valence and arousal in music excerpts. Twenty participants were scanned while listening to 96 music excerpts, classified beforehand into four categories varying in valence and arousal. Music modulated activity in cortical regions, most noticeably in music-specific subregions of the auditory cortex, thalamus, and regions of the reward network such as the amygdala. Using multivoxel pattern analysis, we created a computational model to decode the perceived valence and arousal of the music excerpts with above-chance accuracy. We further explored associations between musical features and brain activity in valence-, arousal-, reward-, and auditory-related networks. The results emphasize the involvement of distinct musical features, notably expressive features such as vibrato and tonal and spectral dissonance in valence, arousal, and reward brain networks. Using ecologically valid music stimuli, we contribute to delineating the neural correlates of music-evoked emotions with potential implications in the development of novel music-based neurorehabilitation strategies.

Index Terms—Music, fMRI, Valence, Arousal, Decoding, Emotion

I. INTRODUCTION

Music can elicit a wide range of affective responses 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 the focus of much research [1]. Understanding the relationship between musical expressivity and affective responses is essential to developing effective and personalized strategies for emotion regulation [2], [3].

Alexandre Sayal is with the Coimbra Institute for Biomedical Imaging and Translational Research (CIBIT), Siemens Healthineers Portugal, and Intelligent Systems Associate Laboratory (LASI). Ana Gabriela Guedes is with CIBIT. Inês Almeida is with CIBIT, LASI, and the Faculty of Medicine of the University of Coimbra (FMUC). Daniela Jardim Pereira is with CIBIT, ULSCoimbra, and FMUC. César F. Lima is with the University Institute of Lisbon (ISCTE). Renato Panda is with the Center for Informatics and Systems of the University of Coimbra (CISUC) and Ci2 Smart Cities Research Center - Polytechnic Institute of Tomar. Rui Pedro Paiva is with CISUC. Teresa Sousa is with CIBIT, the Institute of Nuclear Sciences Applied to Health (ICNAS), FMUC, and LASI. Miguel Castelo-Branco is with CIBIT, ICNAS, FMUC, and LASI. Inês Bernardino is with CIBIT, ICNAS and LASI. Bruno Direito is with CISUC, CIBIT, and LASI.

*alexandresayal@gmail.com

Functional neuroimaging studies have explored the neural bases of music-evoked emotions. Music engages emotion-related brain structures, such as the amygdala, nucleus accumbens (NAcc), hypothalamus, hippocampus, insula, cingulate cortex, and orbitofrontal cortex (OFC) [4]. A meta-analysis [5] showed that music-evoked emotions with positive valence engage the reward network, including the ventral and dorsal striatum (head of the caudate nucleus), amygdala, anterior cingulate cortex (ACC), OFC, insula, mediodorsal thalamus, and secondary somatosensory cortex. Importantly, the recognition of emotions in music may start as early as in the auditory cortex, as [6] postulated, by extending the functional profile of the auditory cortex and hypothesizing an emotion-specific functional hub with connections with a broad range of limbic, paralimbic, and neocortical structures.

Listening to music involves tracking and predicting sound events over time, actions that consistently engage the reward system. Both the perceptual 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 [7]–[10].

Music exploits our expectations through manipulations in melody, rhythm, and their patterns [11]. The relation between specific musical features and the recognized emotions has been studied as a correspondence model based on acoustic features (such as melody, rhythm, harmony, dynamics, or timbre) and formal semantic annotations [12], [13]. To develop technological music-based therapeutic applications (e.g., using brain-computer interfaces [14], [15]), we need to identify the neural correlates of music-evoked emotions and understand how music can convey meaning to the listener [16], [17].

Different theoretical approaches to emotion and music-evoked emotion have been proposed [18], with relevance to emotion generation and regulation [19]. For example, valence and arousal are the fundamental features of the circumplex model of affect [20]. They can be defined as consciously accessible representations of more basic neurophysiological processes, reflecting core descriptive features of all affective experiences [21].

The present study addresses the neural correlates of perceived affective states in music excerpts, as measured by functional magnetic resonance imaging (fMRI). Using multivariate pattern analysis (MVPA), we aim to identify the brain regions that encode musical valence and arousal and explore their relations with music characteristics.

Previous fMRI studies 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 discriminant voxels. Encoding models based on the temporal and frequency dimensions of musical stimuli (musical features encompassing rhythmic, timbral, and tonal properties) revealed a spatial and temporal structure of the underlying neural representations, and decoding performance accuracies were as high as 95% [22]. Clusters in the ACC, insula, and somatosensory cortex were key in the decoding of emotions in music stimuli, particularly joy and fear [23]. Decoding models were also used to identify genres of songs (described by the features 'melody schema' (absolute pitch, relative pitch), and 'acoustic chromagram' (absolute pitch)) considering voxel pattern responses [24].

Here, we aim to identify the brain networks that respond to valence and arousal in music and to understand how acoustic features of the stimuli impact brain activity. Ultimately, we aim to establish the connection between music features, music-derived affective states, and brain activity patterns using MVPA. We hypothesize that the neural networks involved in affective processes [25], 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 previous studies have focused on restricted music stimuli - few genres, music excerpts, or linked to specific emotions such as joy and fear - typically using mass-univariate approaches. Here, we implement a novel approach by employing a diverse set of naturalistic stimuli, thus enhancing the generalizability and robustness of previous findings. We focus on broader affective dimensions by using a two-dimensional system defined in Russell's circumplex model, attempting to establish a combined understanding of both arousal and valence in relation to music perception. By incorporating a wide range of music pieces, each characterized by distinct acoustic features, we are able to identify the neural correlates of these variations. This approach adds to the understanding of how musical features influence brain activity, delving into the complex relationship between music, core affect, and neural processes.

II. METHODS & MATERIALS

A. Participants

Twenty individuals (11 females; mean age 31.3 $\pm$ 6.5 years, from 21 to 41 years) 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 the diagnosis of neurological disorders, mood disorders, or hearing loss.

Participants completed the Profile of Mood States (POMS) questionnaire [26], which includes 42 words describing sensations that people feel in everyday life. They were asked to select an answer, on a scale from 0 (Not at all) to 4 (Very much), based on what best corresponds to the way they have been feeling during the last seven days, including the present day. Mood state was analyzed according to six subscales - Tension (Mdn = 7.5, range: 2-19), Depression (Mdn = 1.5, range: 0-15), Hostility (Mdn = 2.0, range: 2-15), Vigor (Mdn = 13.5, range: 5-22), Fatigue (Mdn = 5.5, range: 0-15), and Confusion (Mdn = 5.0, range: 1-17). A Total Mood Disturbance score revealed a Mdn = 109.5 with a range from 88 to 162.

Our sample reported a mean of 2.2 $\pm$ 3.5 years of formal music training and scored 18.1 $\pm$ 4.6 on average in the Mini-PROMS (Profile of Music Perception Skills) test [27], [28].

B. Music dataset and acoustic analysis

The stimuli were selected from a public dataset - 4Q audio emotion dataset [13] - of 900 30-second audio clips, divided into four categories or quadrants: Q1 (positive valence and high arousal); Q2 (negative valence and high arousal); Q3 (negative valence and low arousal); and Q4 (positive valence and low arousal). The dataset was built using a semi-automatic selection and classification method. Excerpts that were unclear in terms of the perceived valence/arousal were excluded, ending with 225 music excerpts per category. This dataset also includes 1702 features for each clip, both acoustic (such as tempo change and loudness) and musical (such as the number of musical layers, tremolo notes, or vibrato rates). The authors used available audio toolboxes to generate features and created novel, emotionally-relevant audio features based on previous studies. Using statistical classifiers, the authors 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.

C. Procedure

Before being scanned, participants filled out an MRI safety questionnaire and the POMS questionnaire. The fMRI session lasted approximately 75 minutes and was divided into four runs. Participants were asked to perform the listening task with their eyes closed. They listened to 96 musical excerpts randomly selected for each participant from the dataset and balanced across the four quadrants. The excerpts were reduced in duration by selecting the first 11.5 seconds of the original stimulus. Within each run, participants listened to 24 excerpts, 6 of each quadrant, grouped into two sets of 3 stimuli, each lasting 36 seconds (11.5 seconds per excerpt and 0.5 seconds intervals in between). Between music blocks, participants listened to 12 seconds of ambient sound (inherent to the MRI scanner), 12 seconds of white noise, and again 12 seconds of ambient sound (Figure 1). The run structure was based on the pseudo-randomized presentation of music from the four quadrants interleaved with blocks of ambient sound (no auditory stimulus) and of white noise. This randomization ensures that the second presentation of a music block from a quadrant occurs only after the first presentation of all quadrants. The total duration of each run was 10 minutes.
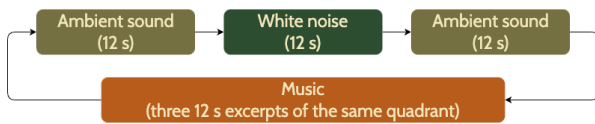
Fig. 1. Diagram of the structure of each trial in the fMRI paradigm. The trials with music from each quadrant were presented twice for each of the four quadrants, always with different excerpts.
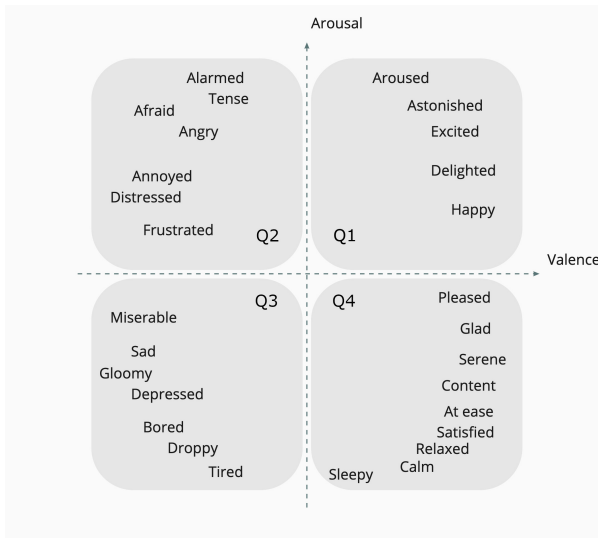

Fig. 2. Arousal-valence plane shown to the participants during the rating experiment. Q1 (positive valence, high arousal), Q2 (negative valence, high arousal), Q3 (negative valence, low arousal), and Q4 (positive valence, low arousal).

After the scanning session, participants listened to the music excerpts again and rated their perceived valence and arousal (the presentation of the music excerpts was randomized per run). The instructions were as follows: "While listening to the music, we ask you to evaluate what it conveys to you. As soon as you think you can classify the musical excerpt in terms of valence and activation, please click with the mouse on the corresponding point in the space, minding that 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). 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.

D. Data acquisition

The MRI session and the rating experiment were carried out using Psychopy version 2022.2.5 [29]. POMS responses were provided by adapting the 42-item paper questionnaire to a digital form.

MRI acquisition was performed on a 3T 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 $1.0 \times 1.0 \times 1.0$ mm, field of view (FOV) = 256×256 mm) and one T2 space sequence (TR = 3200 ms, TE = 408 ms, 192 slices, voxel size $1.0 \times 1.0 \times 1.0$ mm, FOV = 250×250 mm). A total of four functional runs were acquired using a 2D multiband (MB) gradient-echo echo-planar imaging sequence (TR = 1000 ms, TE = 37 ms, flip angle = 68°, 600 volumes, 66 slices, voxel size $2.0 \times 2.0 \times 2.0$ 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 specification [30], using BIDSkit [31] and dcm2niix [32] tools for conversion.

E. Data analysis

We estimated the distribution of valence and arousal ratings provided by the participants for the music excerpts previously heard during the fMRI session. The mouse click position was normalized between -1 and 1 using the furthest report of each participant. We report the match between participants' evaluations and the intended category of the excerpt according to the dataset (Panda, Malheiro, and Paiva 2020).

Imaging data were preprocessed using fMRIPrep 20.2.7 [33], [34], which is based on Nipype 1.7.0 [35], [36]. 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 [37].

1) GLM analysis: The first level analysis was based on a single General Linear Model (GLM) estimated for each participant. The design matrices 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 white noise, corrected for multiple comparisons at the voxel level using False Discovery Rate (FDR) ($q = 0.005$). We extracted 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) Multivoxel decoding analysis: For each participant, a single GLM was re-performed on the preprocessed functional image, in which each trial of each condition of interest was 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 its perceived category, i.e., as identified by the participant in the rating procedure [38].

The model resulted in 96 beta maps (24 music excerpts per run, four runs). The individual beta maps were masked using a gray-matter (GM) mask 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 could 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 [39].

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 while still considering class imbalance.

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 [40].

3) Acoustic features: Previous research investigated the neural processing of individual musical features and generated hypotheses for further studies [41]. In this work, we aimed to comprehensively explore the neural correlates of the acoustic features, particularly in terms of their emotional relevance. We considered the top 100 features defined by [13] as the optimal set to discriminate valence and arousal.

To define the relevance of each feature in the valence, arousal, auditory, and reward networks, we first defined brain masks based on i. large-scale meta-analysis data provided by Neurosynth [42], using the "uniformity-based test" feature (the terms used for the masks were valence, arousal, primary auditory), and ii. cluster 1 from the meta-analysis presented by [5] as a music-specific reward mask. We intersected 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. This method 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 an exploratory one-sample Wilcoxon test using the coefficients obtained (uncorrected).

III. RESULTS

In the following paragraphs, we describe (a) the rating task and compare its results with the original labeling in the music dataset; (b) the neural correlates of music listening following a GLM approach, presenting z-values per condition to characterize valence and arousal in the regions of interest (ROIs); (c) the machine learning approach used to differentiate the four quadrants and present the spatial distribution of the model weights, i.e., the most important voxels for decoding; and (d) the associations between acoustic features and activity in valence, arousal, auditory, and reward networks.

A. Valence and arousal ratings

The results of the rating experiment are summarized in Figure 3. They show some level of mismatch between the labels attributed by the participants and those provided by the 4Q audio emotion dataset. This is particularly the case for the low arousal conditions (Q3 and Q4).

Participants' labels for Q1 (high arousal, positive valence) show that 74.8% of the ratings match the 4Q audio emotion dataset. A lower agreement was detected for Q3 (low arousal, negative valence), where only 39.0% of the ratings matched the 4Q audio emotion dataset, consistent with participant-specific interpretations of the emotional content of the excerpts. Because of these differences, all the following analyses were based on the participants' labels.

Figure 4 shows the distribution of participants' arousal and valence ratings for each quadrant, with quadrants defined according to the dataset labeling. The results show that Q3 stimuli tended to be perceived as neutral here, not negative as expected. For arousal, the pattern was generally consistent with the prior labeling.

B. Neural correlates of 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, thalamus, cerebellum, and motor cortices, as well as in regions of the reward network such as the amygdala and putamen. Table I presents the full list of clusters, and Figure 6 the z-values for each quadrant in each cluster. We found differences among quadrants in the

Fig. 3. Confusion matrix comparing the labels originally associated with the music clips in the database with those attributed by the participants in the current study. 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.

TABLE I
SIGNIFICANT CLUSTERS FOR THE CONTRAST BETWEEN 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	Size (mm^3)	Labels
1	58	-5	2	7.29		
1a	50	-15	4	7.05	34968	Temporal superior sulcus, auditory cortex
1b	66	-29	8	6.89		
1c	44	-25	6	6.82		
2	-59	-11	4	7.01		
2a	-49	-17	2	6.95	34376	Temporal superior sulcus, auditory cortex
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		
5a	30	-61	-47	4.89	2376	Cerebellum (VI)
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		
7a	-15	-69	-53	4.54	2416	Cerebellum (vermis, VI)
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

auditory cortices, thalamus, cerebellum, and supplementary motor area (SMA), with high arousal and positive music (Q1) eliciting more activity in the majority of the clusters. Additionally, we provide the z-scores contrasting positive and negative arousal and valence in supplementary Figures S14 and S15, respectively.

C. Predicting perceived valence and arousal from fMRI data

In the following classification analyses, we considered data from 15 participants. As the classes were defined according to each participant's rating, and four participants did not report ratings for one of the quadrants, their data were not included.

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 stringent dummy predictor is statistically significant ($p < 0.0001$).

2) Visualization of weights: To explore the spatial distribution of the voxels that better differentiated quadrants, we selected the top 2% most discriminant voxels for each one vs. all models (Q1 vs. others, Q2 vs. others, and so on) and overlapped them. 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 for all quadrants. The full list of clusters can be found in Table II.

D. Acoustic features and brain activation

To explore the impact of acoustic features in the valence, arousal, auditory, and reward networks, we investigated its regression profile with all acoustic features, by computing the mean over all voxels of each mask per participant.

In this analysis, we included a set of 100 candidate acoustic features based on the previous work and nomenclature by [13].

For a complete overview of these analyses, see Supplementary Materials.

For the exploratory analysis considering the valence masks, four of the seven acoustic features presenting significant results were expressivity features. In particular, these features are based on vibrato, an expressive technique used in vocal and instrumental music that consists of a regular pitch oscillation. The main vibrato-related features are the amount of pitch variation (extent) and the velocity (rate) of this pitch variation. Similarly, we see a dominance of expressivity features in the arousal mask. Regarding the auditory mask, 54 features appeared to be relevant, with most being 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.
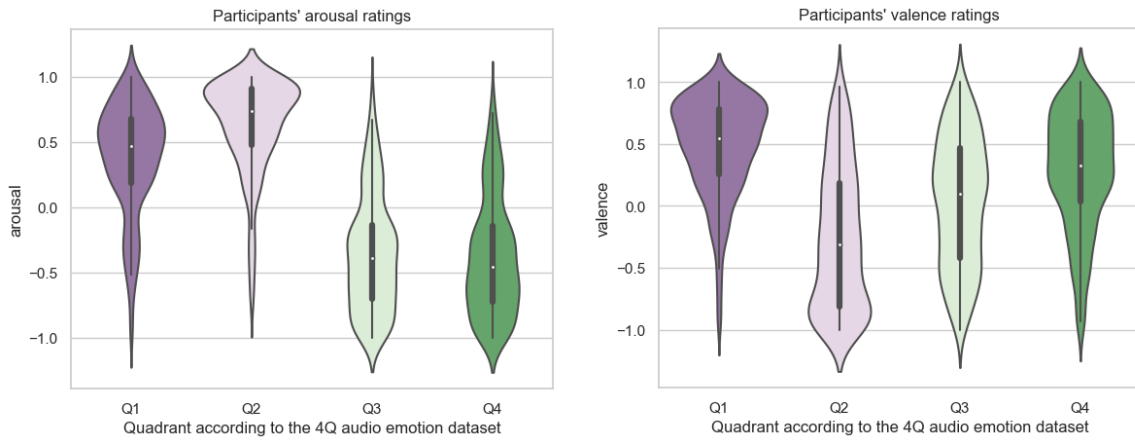
Fig. 4. Violin plots displaying the symmetric kernel density estimate and the quartiles of a box plot of the participants' valence and arousal ratings for each music category (as categorized in the original dataset). The values displayed are the Euclidean distance from each axis to the position in the plane where the participants clicked, normalized for the furthest report.
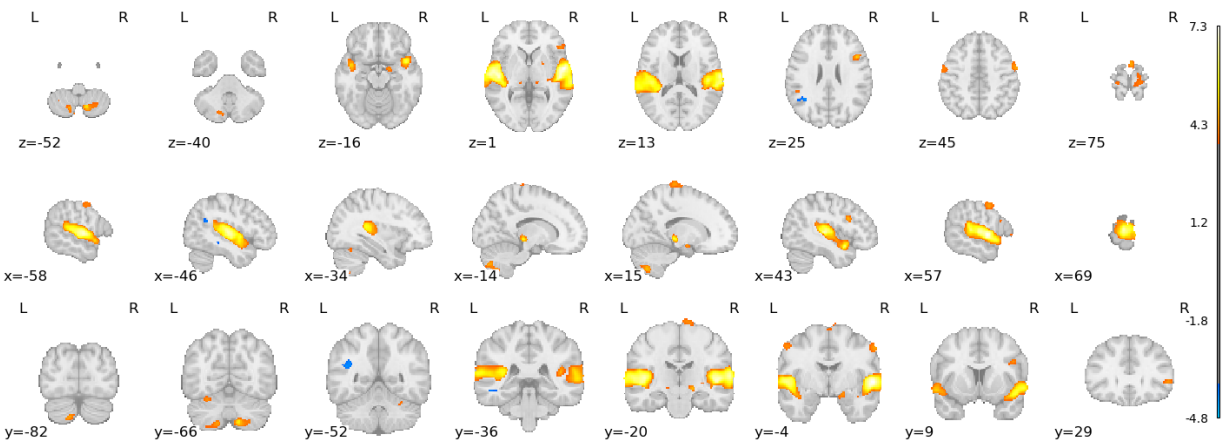

Fig. 5. Group activation map for the contrast between all quadrants and noise (z-values FDR corrected, $q = 0.005$, cluster correction k < 25, N = 19).
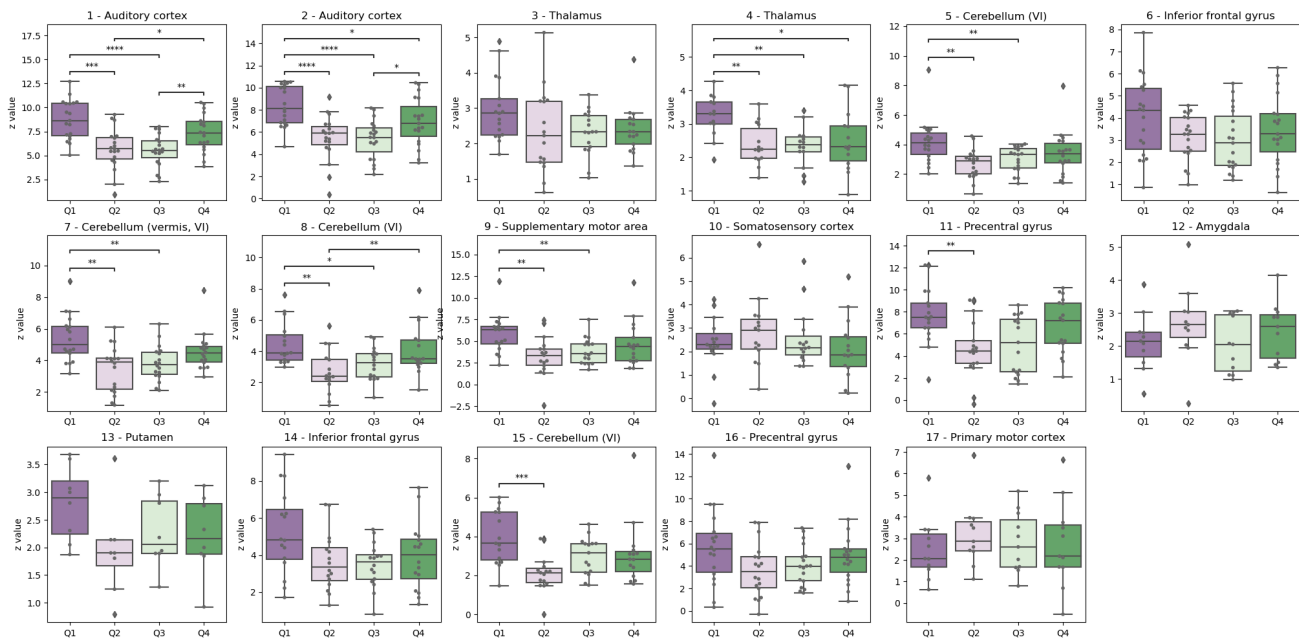

Fig. 6. Boxplots for the z-values of each identified cluster (Table I) per quadrant. Pairwise statistical comparisons between quadrants are displayed, performed using two-sided Mann-Whitney-Wilcoxon tests with FDR correction ($q = 0.05$).
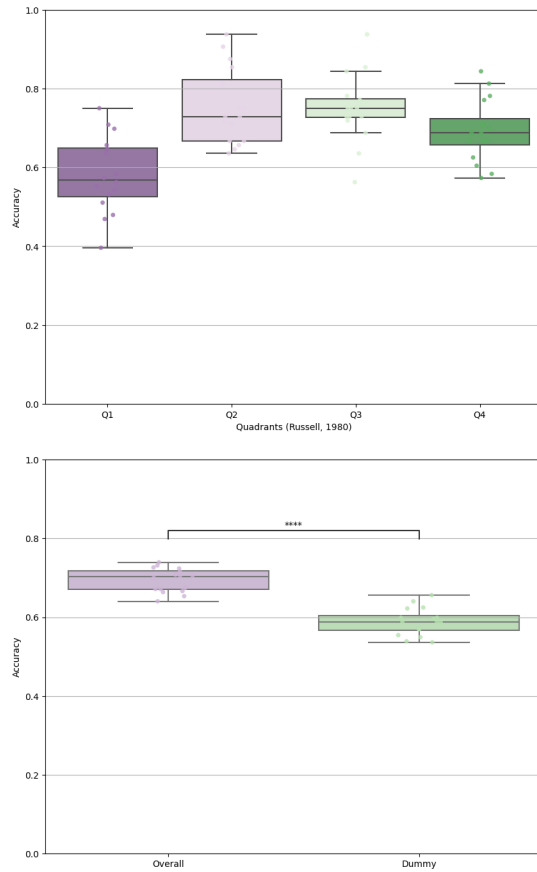
Fig. 7. MVPA classification accuracy. On the top, we display the boxplot of the accuracies per quadrant. On the bottom, we compare the mean overall accuracy with the accuracy of the dummy predictor (**** - $p < 0.0001$).

TABLE II
CLUSTERS IDENTIFIED IN THE MAP OF FIGURE 8, DISPLAYING THE COORDINATES OF THE PEAK VOXEL, THE NUMBER OF QUADRANTS AT THAT COORDINATE, THE CLUSTER SIZE, AND THE LABELS FROM THE AAL3 ATLAS AND NEUROSYNTH ASSOCIATIONS.

X	Y	Z	#Q at peak	Size (mm^3)	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, prefrontal cortex, working memory
44	6	40	1	800	Frontal middle gyrus, prefrontal cortex
56	-1	46	1	816	Precentral gyrus, motor

IV. DISCUSSION

Based on a dataset including diverse music genres, we found that brain activity during music listening predicted perceived valence and arousal with above-chance accuracy. This classification was mainly based on bilateral auditory cortices, with additional contributions from the precuneus, frontal and precentral gyri, and cingulate cortex.

Furthermore, the GLM analysis showed that music listening 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 SMA.

The acoustic features that significantly predicted brain activity differed across the valence, arousal, auditory, and reward networks. This analysis highlighted expressive features, particularly the vibrato-related ones, as network modulators.

These findings contribute to further comprehending how music relates to emotions, defining music-specific regions of interest, and identifying musical features that modulate different brain networks. Ultimately, we pave the way for developing music-based neurorehabilitative strategies (such as neurofeedback): the inherent ability of music to induce and regulate emotions and 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 [43].

A. Valence and Arousal ratings

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 [44]. Schubert defined the former as the internal locus of emotion and the latter as the external locus and discussed the relation between them [2]. The wording used in our rating task emphasized the external locus. The short duration of the music excerpts (aimed to explore a wide range of genres and acoustic features) was more suited to study perceived rather than felt emotions. Nevertheless, [23] have suggested that feelings could emerge within seconds, suggesting the possibility of using shorter musical stimuli.

Three main theories for the distinction and interaction of felt and perceived emotions were proposed: emotion contagion (both loci are part of the same emotional process, which explains why we tend to feel the emotion that music is conveying), 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 felt emotions are often rated the same (particularly in the arousal factor) or lower than the corresponding emotion expressed by the stimuli [2]. By using a dataset with preselected emotion labels, we aimed to balance the stimuli on arousal and valence. Nevertheless, the observed differences between predefined labels and participants' categorizations highlight the need to tailor the task individually. Several
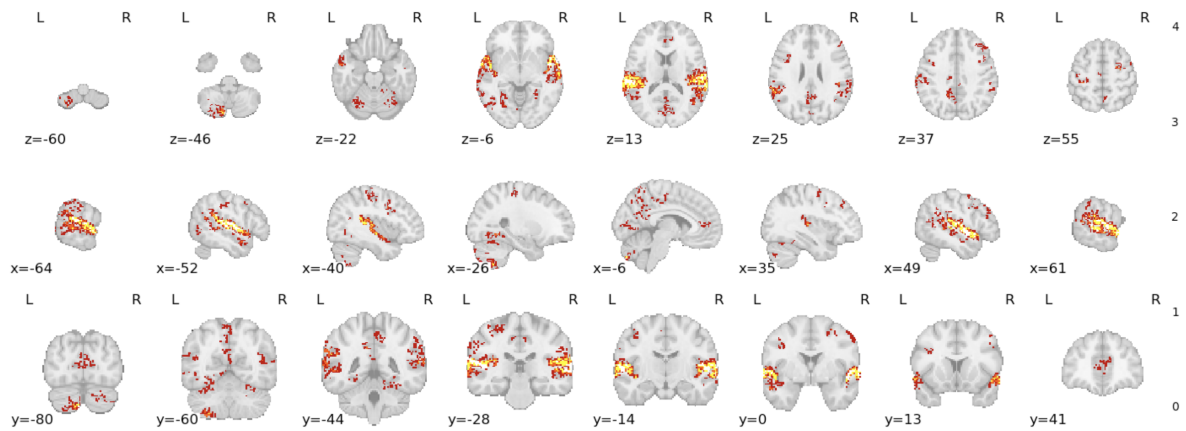
Fig. 8. The combination of the top 2% most important voxels for each quadrant to classify valence and arousal categories in the whole-brain analysis (the colorbar represents the number of times that a specific voxel appears in the top 2%).

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, extensively described in the music literature and interpreted as a paradox of tragedy (deriving pleasure from something that is tragic), may also contribute to the shift of the ratings from negative to positive [45]–[47]. All the analyses presented in this study considered the labels provided by each participant at the cost of stimuli imbalance between classes.

Low arousal stimuli presented a significant mismatch with the labels provided in the dataset. The ambiguity associated with low arousal has been previously described for music [48] and for other stimuli such as facial expressions [49]. 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 [12], [50].

B. Neural correlates of music-induced valence and arousal

1) GLM and activation per quadrant: The contrast of music vs. noise revealed 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 multisensory integration) and thus plays a role in music listening and production [4]. Because the control condition was noise, at least a component of the activations was not due to the mere presence of sound. In accordance with previous findings, the auditory cortex contributes not only to sensory perception but also to higher-level aspects of auditory processing. Responses in bilateral auditory cortices varied as a function of valence, as pieces of music labeled as Q1 and Q4 elicited stronger responses than those labeled as Q2 and Q3 (i.e., activity was stronger for positive vs. negative music). Recent studies have shown that emotional/motivational

processes influence activity in the auditory cortex and that neurons in the auditory cortex can distinguish between positive and negative sounds [51]. One of the models proposed for this emotion-informed process in sensory cortices is based on evidence that the reward value of music is correlated with increasing functional connectivity inside the music perception network [5].

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 [52]. This structure is also involved in emotional reactions to acoustic stimuli and is part of a thalamocortical circuit that integrates musical context and content [53].

Frontal activity was found in the inferior frontal gyrus (pars triangularis and frontal operculum), particularly in the right hemisphere. [54] also reported increased activity in the inferior frontal gyrus when 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 [55]; fear and joy in [23]).

Clusters were also 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 in auditory processing, emotion perception, and the subjective experience of emotions [4], [56]. Notably, pleasurable music activates motor regions, including the SMA and cerebellum, more strongly than music that is emotionally less evocative (note that brain activity during pieces of music labeled as Q1 presented the highest values) [57]. Music frequently prompts spontaneous rhythmic movements in listeners. Therefore, the sensorimotor activity observed in the present study may reflect the interplay between perception and movement, even without overt physical actions.

The GLM also revealed activity in the amygdala during music listening. Moreover, the activation pattern in the amygdala

did not show significant differences associated with valence or arousal of the music pieces. Koelsch and colleagues [58] found that brain activity in the amygdala was higher for joy compared to fear. Our finding is more consistent with the view that the amygdala is an emotional processing hub, important for both negative and positive stimuli [59].

The hippocampus has been previously linked to music-evoked emotions. Besides the hippocampal-auditory connections, which are key to long-term auditive memory, evidence suggests that this region plays a role in emotional function and in the formation and maintenance of social attachments [53], [60], [61]. Moreover, its connections to reward structures further implicate it in the emotional processing of music [62].

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 substantia 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 [63].

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 [64]. Different parts of the cerebellum and its connectivity with the cerebral cortex and the basal ganglia have been associated with emotion identification and pleasure in music-listening tasks. It has been shown, for example, that the cerebellum is involved in processing acoustic features such as timbre and pitch [24], [65].

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 decoding analysis reinforces the role of structures such as the bilateral auditory cortices, auditory thalamus, SMA, and cerebellum in decoding valence and arousal.

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. The results suggest a significant overlap between the main regions identified here and the GLM analysis.

C. How does music express different emotions? The neural correlates of acoustic features

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 related to discrete emotions [41].

Regarding the reward mask, we found that expressivity, loudness (tonal and spectral dissonance), and pitch were 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 [11], [60], [65].

[66] studied physical voice features and concluded that the amygdala and auditory cortex decode the affective quality of a voice by receiving inputs from other brain regions that process acoustic features, while also directly responding to specific acoustic features themselves without requiring prior processing, particularly when these features are critical for determining the emotional value of the voice.

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" [67]. In addition to being a characteristic of the singing voice [68], vibrato is also used in most musical instruments for expressivity. Some even argue that vibrato emerged because it mimics the crying voice, and the manipulation of its extent, rate, or intensity may be linked to/identified as different emotions [69].

[22] 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 such 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. [70] 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 STG, 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. [24] 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 expressive 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 [53], [71]. Therefore, our data suggests that musical features are directly or indirectly linked to the activation patterns of different brain areas linked

to valence and arousal [72].

D. Limitations

We used a public dataset as stimuli to obtain a diverse yet balanced dataset in valence and arousal. 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 reduced the power of our analysis. This limitation is minimized by the ability of models to adjust weights and decision hyperplanes per class.

To maximize the heterogeneity and variety of music pieces in the dataset, we selected a relatively short segment of each piece. Nevertheless, recent evidence [23] has 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.

E. Conclusion

Our results indicate that several brain regions significantly encode perceived valence and arousal during listening to naturalistic music. Using a comprehensive dataset, we probed the neural correlates of valence and arousal in music. We also explored how several acoustic features modulated activity in key networks associated with valence, arousal, sensory processing, and reward.

Although our emphasis was on perceived rather than felt emotions, the emotion contagion theory highlights the close link between the two processes. Nevertheless, future studies targeting felt emotions, or neurorehabilitative approaches targeting emotion and mood regulation, should take the optimization of the emotion framework into consideration. Previous studies indicate that familiar music elicits stronger emotional reactions and engagement of reward structures [57]. 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. These studies could also use longer musical segments and self-selected musical pieces to optimize the induction of emotional reactions. Moreover, focusing on valence and arousal alone may not be optimal for exploring felt emotions. Aesthetic emotions [73], for example, may represent a more appropriate model. The connectivity patterns involved in musical emotions should also be addressed in future studies. [8] 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 associations between music acoustic features and brain activity (e.g. in the reward system) is key to optimizing therapeutic strategies. Brain-computer interfaces (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, as regulators of volitional and non-volitional brain activity. Real-time adjustment of specific music features may represent an important alternative for emotion regulation of BCIs [74]. Our results suggest that other acoustic features (for example vibrato) may ultimately represent a better choice to regulate brain activity linked to valence and reward.

DATA AVAILABILITY

All scripts used for the analyses mentioned in this work can be found at https://github.com/CIBIT-UC/brainplayback_task01. The anonymized and defaced dataset can be found in BIDS format at https://doi.org/10.57979/OM4UWT. The data management plan regarding this project can be found at https://doi.org/10.5281/zenodo.10563830.

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, Dec. 2012. [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.

[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.

[4] 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, p. 20813, Oct. 2021. [Online]. Available: https://www.nature.com/articles/s41598-021-00139-3

[5] S. Koelsch, "A coordinate-based meta-analysis of music-evoked emotions," *NeuroImage*, vol. 223, p. 117350, Dec. 2020. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1053811920308363

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

[7] B. P. Gold, M. T. Pearce, E. Mas-Herrero *et al.*, "Predictability and Uncertainty in the Pleasure of Music: A Reward for Learning?" *The Journal of Neuroscience*, vol. 39, no. 47, pp. 9397–9409, Nov. 2019. [Online]. Available: https://www.jneurosci.org/lookup/doi/10.1523/JNEUROSCI.0428-19.2019

[8] V. N. Salimpoor, D. H. Zald, R. J. Zatorre *et al.*, "Predictions and the brain: How musical sounds become rewarding," *Trends in Cognitive Sciences*, vol. 19, no. 2, pp. 86–91, 2015, publisher: Elsevier Ltd. [Online]. Available: http://dx.doi.org/10.1016/j.tics.2014.12.001

[9] 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.

[10] D. Huron, *Sweet Anticipation: Music and the Psychology of Expectation*. The MIT Press, 2006. [Online]. Available: https://direct.mit.edu/books/book/1961/sweet-anticipationmusic-and-the-psychology-of

[11] H. Chapin, K. Jantzen, J. A. Scott Kelso *et al.*, "Dynamic Emotional and Neural Responses to Music Depend on Performance Expression and Listener Experience," *PLoS ONE*, vol. 5, no. 12, p. e13812, Dec. 2010. [Online]. Available: https://dx.plos.org/10.1371/journal.pone.0013812

[12] J. S. Gomez-Canon, E. Cano, T. Eerola *et al.*, "Music Emotion Recognition: Toward new, robust standards in personalized and context-sensitive applications," *IEEE Signal Processing Magazine*, vol. 38, no. 6, pp. 106–114, Nov. 2021. [Online]. Available: https://ieeexplore.ieee.org/document/9591555/

[13] 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, Oct. 2020. [Online]. Available: https://ieeexplore.ieee.org/document/8327886/

[14] R. Ramirez, M. Palencia-Lefler, S. Giraldo, and Z. Vamvakousis, "Musical neurofeedback for treating depression in elderly people," *Frontiers in Neuroscience*, vol. 9, Oct. 2015. [Online]. Available: http://journal.frontiersin.org/Article/10.3389/fnins.2015.00354/abstract

[15] W. Trost, C. Trevor, N. Fernandez *et al.*, "Live music stimulates the affective brain and emotionally entrains listeners in real time," *Proceedings of the National Academy of Sciences*, vol. 121, no. 10, p. e2316306121, Mar. 2024. [Online]. Available: https://pnas.org/doi/10.1073/pnas.2316306121

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

[17] 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, Oct. 2008. [Online]. Available: https://www.cambridge.org/core/product/identifier/S0140525X08005293/type/journal_article

[18] L. A. Warrenburg, "Comparing musical and psychological emotion theories." *Psychomusicology: Music, Mind, and Brain*, vol. 30, no. 1, pp. 1–19, Mar. 2020. [Online]. Available: https://doi.apa.org/doi/10.1037/pmu0000247

[19] 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, Jan. 2011. [Online]. Available: http://journals.sagepub.com/doi/10.1177/1754073910380974

[20] J. A. Russell, "A circumplex model of affect," *Journal of Personality and Social Psychology*, vol. 39, no. 6, pp. 1161–1178, 1980.

[21] 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, Jan. 2016.

[22] S. Hoefle, A. Engel, R. Basilio *et al.*, "Identifying musical pieces from fMRI data using encoding and decoding models," *Scientific Reports*, vol. 8, no. 1, pp. 1–13, 2018.

[23] 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. [Online]. Available: https://doi.org/10.1038/s41598-021-89405-y

[24] 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.

[25] L. F. Barrett and E. Bliss-Moreau, "Chapter 4 Affect as a Psychological Primitive," in *Advances in Experimental Social Psychology*. Elsevier, 2009, vol. 41, pp. 167–218. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0065260108004048

[26] 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. [Online]. Available: http://publicacoes.ispa.pt/index.php/ap/article/view/345

[27] L. N. C. Law and M. Zentner, "Assessing Musical Abilities Objectively: Construction and Validation of the Profile of Music Perception Skills," *PLoS ONE*, vol. 7, no. 12, p. e52508, Dec. 2012. [Online]. Available: https://dx.plos.org/10.1371/journal.pone.0052508

[28] M. Zentner and H. Strauss, "Assessing musical ability quickly and objectively: development and validation of the Short-PROMS and the Mini-PROMS," *Annals of the New York Academy of Sciences*, vol. 1400, no. 1, pp. 33–45, Jul. 2017. [Online]. Available: https://nyaspubs.onlinelibrary.wiley.com/doi/10.1111/nyas.13410

[29] J. Peirce, J. R. Gray, S. Simpson *et al.*, "PsychoPy2: Experiments in behavior made easy," *Behavior Research Methods*, vol. 51, no. 1, pp. 195–203, Feb. 2019. [Online]. Available: http://link.springer.com/10.3758/s13428-018-01193-y

[30] K. J. Gorgolewski, T. Auer, V. D. Calhoun *et al.*, "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.

[31] M. Tyszka, "bidskit," 2023. [Online]. Available: https://github.com/jmtyszka/bidskit

[32] X. Li, P. S. Morgan, J. Ashburner *et al.*, "The first step for neuroimaging data analysis: DICOM to NIfTI conversion," *Journal of Neuroscience Methods*, vol. 264, pp. 47–56, May 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0165027016300073

[33] O. Esteban, C. J. Markiewicz, R. W. Blair *et al.*, "fMRIPrep: a robust preprocessing pipeline for functional MRI," *Nature Methods*, vol. 16, no. 1, pp. 111–116, Jan. 2019. [Online]. Available: https://www.nature.com/articles/s41592-018-0235-4

[34] O. Esteban, C. J. Markiewicz, M. Goncalves *et al.*, "fMRIPrep: a robust preprocessing pipeline for functional MRI," Aug. 2023. [Online]. Available: https://zenodo.org/record/852659

[35] K. Gorgolewski, C. D. Burns, C. Madison *et al.*, "Nipype: A Flexible, Lightweight and Extensible Neuroimaging Data Processing Framework in Python," *Frontiers in Neuroinformatics*, vol. 5, 2011. [Online]. Available: http://journal.frontiersin.org/article/10.3389/fninf.2011.00013/abstract

[36] O. Esteban, C. J. Markiewicz, C. Burns *et al.*, "nipy/nipype: 1.8.3," Jul. 2022. [Online]. Available: https://zenodo.org/record/596855

[37] A. Abraham, F. Pedregosa, M. Eickenberg *et al.*, "Machine learning for neuroimaging with scikit-learn," *Frontiers in Neuroinformatics*, vol. 8, 2014. [Online]. Available: http://journal.frontiersin.org/article/10.3389/fninf.2014.00014/abstract

[38] 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. [Online]. Available: http://dx.doi.org/10.1016/j.neuroimage.2011.08.076

[39] G. Varoquaux, P. R. Raamana, D. A. Engemann *et al.*, "Assessing and tuning brain decoders: Cross-validation, caveats, and guidelines," *NeuroImage*, vol. 145, pp. 166–179, Jan. 2017. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S105381191630595X

[40] E. T. Rolls, C.-C. Huang, C.-P. Lin *et al.*, "Automated anatomical labelling atlas 3," *NeuroImage*, vol. 206, p. 116189, Feb. 2020. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1053811919307803

[41] 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, Sep. 2004. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/0929821042000317813

[42] T. Yarkoni, R. A. Poldrack, T. E. Nichols *et al.*, "Large-scale automated synthesis of human functional neuroimaging data," *Nature Methods*, vol. 8, no. 8, pp. 665–670, 2011, iSBN: 1662-5196.

[43] B. Direito, M. Ramos, J. Pereira *et al.*, "Directly exploring the neural correlates of feedback-related reward saliency during fMRI- based Neurofeedback," *Frontiers in Human Neuroscience*, vol. 14, no. February, p. 578119, 2021, arXiv: 1011.1669v3 ISBN: 9788578110796.

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

[45] T. Eerola, J. K. Vuoskoski, H.-R. Peltola *et al.*, "An integrative review of the enjoyment of sadness associated with music," *Physics of Life Reviews*, vol. 25, pp. 100–121, Aug. 2018. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S157106451730163X

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

[47] 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, Mar. 2023. [Online]. Available: https://bmcpsychology.biomedcentral.com/articles/10.1186/s40359-023-01110-9

[48] 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, Apr. 2017. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=18563

[49] A. J. Gerber, J. Posner, D. Gorman *et al.*, "An Affective Circumplex Model of Neural Systems Subserving Valence, Arousal, & Cognitive Overlay During the Appraisal of Emotional Faces," *Neuropsychologia*, 2009.

[50] 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, May 2008. [Online]. Available: http://doi.apa.org/getdoi.cfm?doi=10.1037/1931-3896.2.2.115

[51] 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, Mar. 2019. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0149763418307619

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

[53] 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, Jul. 2014. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1002/hbm.22416

[54] 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. [Online]. Available: http://www.hindawi.com/journals/bn/2015/529043/

[55] V. Putkinen, S. Nazari-Farsani, K. Seppälä *et al.*, "Decoding Music-Evoked Emotions in the Auditory and Motor Cortex," *Cerebral Cortex*, vol. 31, no. 5, pp. 2549–2560, Mar. 2021. [Online]. Available: https://academic.oup.com/cercor/article/31/5/2549/6046263

[56] 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. [Online]. Available: http://dx.doi.org/10.1016/j.tins.2016.06.003

[57] 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, Sep. 2001. [Online]. Available: https://pnas.org/doi/full/10.1073/pnas.191355898

[58] S. Koelsch, S. Skouras, T. Fritz *et al.*, "The roles of superficial amygdala and auditory cortex in music-evoked fear and joy," *NeuroImage*, vol. 81, pp. 49–60, 2013, publisher: Elsevier Inc. [Online]. Available: http://dx.doi.org/10.1016/j.neuroimage.2013.05.008

[59] K. A. Lindquist, A. B. Satpute, T. D. Wager *et al.*, "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, May 2016. [Online]. Available: https://academic.oup.com/cercor/article-lookup/doi/10.1093/cercor/bhv001

[60] S. Koelsch, T. Fritz, D. Y. V. Cramon *et al.*, "Investigating emotion with music: An fMRI study," *Human Brain Mapping*, vol. 27, no. 3, pp. 239–250, Mar. 2006. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1002/hbm.20180

[61] R. W. Wilkins, D. A. Hodges, P. J. Laurienti *et al.*, "Network science and the effects of music preference on functional Brain connectivity: From Beethoven to Eminem," *Scientific Reports*, vol. 4, pp. 1–8, 2014.

[62] H.-E. Schaefer, "Music-Evoked Emotions—Current Studies," *Frontiers in Neuroscience*, vol. 11, p. 600, Nov. 2017. [Online]. Available: http://journal.frontiersin.org/article/10.3389/fnins.2017.00600/full

[63] L. Ferreri, E. Mas-Herrero, R. J. Zatorre *et al.*, "Dopamine modulates the reward experiences elicited by music," *Proceedings of the National Academy of Sciences*, vol. 116, no. 9, pp. 3793–3798, Feb. 2019. [Online]. Available: https://pnas.org/doi/full/10.1073/pnas.1811878116

[64] 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. [Online]. Available: https://www.nature.com/articles/nrn2152

[65] V. Alluri, P. Toiviainen, I. P. Jääskeläinen *et al.*, "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. [Online]. Available: http://dx.doi.org/10.1016/j.neuroimage.2011.11.019

[66] 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, Dec. 2016. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0010945216303045

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

[68] 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, Mar. 2015. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S089219971400126X

[69] F. Eyben, G. L. Salomão, J. Sundberg *et al.*, "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, Dec. 2015. [Online]. Available: https://asmp-eurasipjournals.springeropen.com/articles/10.1186/s13636-015-0057-6

[70] P. Toiviainen, V. Alluri, E. Brattico *et al.*, "Capturing the musical brain with Lasso: Dynamic decoding of musical features from fMRI data," *NeuroImage*, vol. 88, pp. 170–180, 2014, publisher: Elsevier Inc. [Online]. Available: http://dx.doi.org/10.1016/j.neuroimage.2013.11.017

[71] 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.

[72] S.-G. Kim, K. Mueller, J. Lepsien *et al.*, "Brain networks underlying aesthetic appreciation as modulated by interaction of the spectral and temporal organisations of music," *Scientific Reports*, vol. 9, no. 1, p. 19446, Dec. 2019. [Online]. Available: https://www.nature.com/articles/s41598-019-55781-9

[73] 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. [Online]. Available: http://doi.apa.org/getdoi.cfm?doi=10.1037/1528-3542.8.4.494

[74] K. Takabatake, N. Kunii, H. Nakatomi *et al.*, "Musical Auditory Alpha Wave Neurofeedback: Validation and Cognitive Perspectives," *Applied Psychophysiology and Biofeedback*, vol. 46, no. 4, pp. 323–334, Dec. 2021. [Online]. Available: https://link.springer.com/10.1007/s10484-021-09507-1

Alexandre Sayal completed the MSc in Biomedical Engineering in 2016 at the University of Coimbra, Portugal. He has been researching in the field of cognitive neuroscience with a particular interest in functional magnetic resonance imaging. Three years as a research fellow at the Coimbra Institute for Biomedical Imaging and Translational Research (CIBIT) led him to enroll in the Biomedical Engineering PhD at the same university in 2019. The project combines his passion for music with an engineer's will to further understand the most complex biological structure - the human brain.

Ana Gabriela Guedes completed her Master's degree in Biomedical Engineering at the University of Coimbra in 2023, specializing in imaging and radiation. During 2023, she collaborated with the Coimbra Institute for Biomedical Imaging and Translational Research (CIBIT) on projects involving machine learning and brain imaging, as well as the study of visual hallucinations. Since 2024, she has worked as a biomedical engineer at Critical Software.

Inês Almeida has a degree in Psychology (2006), and a PhD in Health Sciences (2014), from the University of Coimbra. She has worked in the interconnection between Psychology and Neuroscience, using psychological assessment, psychophysiology, neuroimaging, and human behavior measures to study memory, decision-making, social cognition, and affective processes in healthy and clinical populations. She has recently started a junior researcher position in the meta-research ERA Chair project EXCELScIOR at the Centre for Innovative Biomedicine and Biotechnology (CIBB), University of Coimbra.

Daniela Jardim Pereira is a neuroradiologist at ULS Coimbra, a guest researcher at the Coimbra Institute for Biomedical Imaging and Translational Research (CIBIT), and an invited Assistant Professor at the Faculty of Medicine, University of Coimbra. She earned her medical degree in 2010 and completed her PhD in 2023. With a multidisciplinary background, her expertise spans structural and functional neuroimaging, cognitive neuroscience, and neuroanatomy, with a focus on both healthy and pathological brain states.

César F. Lima is an Associate Professor of Psychology at the University Institute of Lisbon (ISCTE). His research focuses on the psychology and neuroscience of music, voice, and emotion. He investigates these topics across diverse populations, including healthy individuals of all ages, specialized groups such as musicians, and clinical populations. César F. Lima completed his PhD in Psychology at the University of Porto (2011) and was a Postdoctoral Research Associate at the Institute of Cognitive Neuroscience - University College London (2012-2016).

Renato Panda is an Auxiliary Researcher at Ci2, Polytechnic Institute of Tomar. His main research interests are Music Emotion Recognition (MER) and Music Information Retrieval (MIR). He earned his PhD in Informatics Engineering from the University of Coimbra in 2019. Since then, he has been part of the Cognitive and Media Systems group at the Centre for Informatics and Systems of the University of Coimbra (CISUC), where he remains actively involved.

Rui Pedro Paiva is a Professor at the Department of Informatics Engineering of the University of Coimbra, where he concluded his Doctoral, Master, and Bachelor degrees in 2007, 1999, and 1996, respectively. He is also a member of the CMS group at CISUC. His main research interests are in the areas of Music Information Retrieval and Health Informatics. The common research hat is the study of feature engineering, machine learning, and signal processing to analyze musical and biosignals.

Teresa Sousa is an Auxiliary Researcher in biosignals and neurorehabilitation at the Faculty of Medicine of the University of Coimbra. Her research interests include various neuromodulation approaches for treatment-resistant neuropsychiatric conditions and neurodevelopmental disorders, as well as advanced brain mapping to better characterize brain function and therapeutic response heterogeneity. She graduated in Biomedical Engineering from the University of Minho in 2010 and completed her PhD in the same field at the University of Coimbra in 2017.

Miguel Castelo-Branco obtained his PhD at the Max-Planck Institute for Brain Research, Germany, held a Professorship at the University of Maastricht, the Netherlands, and is now Full Professor at the University of Coimbra (UC). He made interdisciplinary contributions in the fields of Brain-Computer Interfaces, Cognitive Neuroscience, Human and Animal Neurophysiology, Visual Neuroscience, Human Psychophysics, Functional Brain Imaging and translational research in Neuropsychiatric disorders. He is the Scientific Coordinator of the National Functional Brain Imaging Network. He is the Director of CIBIT, rated as Excellent by International evaluation, and was the Director of ICNAS, Medical Imaging Infrastructure (UC).

Inês Bernardino is an auxiliary researcher at CIBIT-ICNAS, University of Coimbra. She graduated in Psychology from the University of Coimbra in 2007 and completed a PhD in Health Sciences from the Faculty of Medicine of the University of Coimbra in 2014. She has built up a multidisciplinary background in Cognitive Neuroscience, Biomedical Sciences, and Neuropsychology. Her expertise lies in developing and implementing cognitive and imaging studies in clinical populations employing molecular, physiological, and behavioral approaches.

Bruno Direito earned his PhD in Information Sciences and Technologies from the University of Coimbra in 2013. From 2013 to 2021, he was a research fellow at the Faculty of Medicine, University of Coimbra. Currently, he is an Auxiliary Researcher with the Adaptive Computing Group at the Centre for Informatics and Systems, University of Coimbra. His research focuses on biomedical signal processing, pattern recognition, and the development of rehabilitative applications using signal-processing techniques.