UNIVERSIDADE D
COIMBRA

Ana Gabriela de Almeida Guedes

Music of the brain:
A pattern recognition framework to investigate the neural correlates of music

Thesis submitted to the Faculty of Sciences and Technology of the University of Coimbra for the degree of Master in Biomedical Engineering with specialization in Imaging and Radiation, supervised by Prof. Dr. Bruno Direito

September 2023

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE D COIMBRA

Music of the brain:

A pattern recognition framework to investigate the neural correlates of music

Ana Gabriela de Almeida Guedes

Supervisor:
Doutor Bruno Direito

Thesis submitted to the Faculty of Sciences and Technology of the University of Coimbra to obtain a degree of master in biomedical engineering

September 2023

This work was developed in collaboration with:

Instituto de Ciências Nucleares Aplicadas à Saúde da Universidade de Coimbra

Coimbra Institute for Biomedical Imaging and Translational Research

Funded by:

FCT project expl/psi-ger/0948/2021

Agradecimentos

Concluído este trabalho, não posso deixar de agradecer a um conjunto de pessoas sem as quais este não teria sido possível:

Em primeiro lugar, ao meu orientador, Doutor Bruno Direito, por todo o acompanhamento e ensinamentos imprescindíveis ao longo destes meses. Pela calma e paciência a explicar mesmo as coisas mais simples, e por salientar sempre que as únicas perguntas parvas são as que ficam por colocar. Foi um privilégio aprender e trabalhar sob a sua orientação.

Em segundo, talvez o agradecimento mais importante que posso fazer: à minha família. Aos meus pais, que são o meu maior exemplo a todos os níveis e me ensinaram o significado de amor e sacrifício. Por serem o meu ninho de calma no meio do caos e pelo relembrar constante de que descanso também é preciso e que tudo se faz. Por serem colo, compreensão e motivação incessantes, obrigada. À metade mais diferente de mim, Salomé, por equilibrares sempre a balança dos sorrisos e lágrimas e estares comigo em cada minuto deste processo. Se aqui apresento este trabalho, ele deve-se quase tanto a ti como a mim. Ao meu irmão, que me orgulha dia após dia com a sua capacidade de foco e dedicação e o seu coração bonito, pelo interesse genuíno que demonstrou desde o dia 1 em querer perceber o que eu andava realmente a fazer. Aos meus avós, por me relembrarem sempre que "dos fracos não reza a História" e me ensinarem a lidar com as adversidades da vida de sorriso na cara e assobio nos lábios. Aos meus tios e padrinhos, pelo apoio incondicional de todas as formas. E por fim, à minha Mila Camila, fiel companheira da vida. A todos vocês, obrigada por acreditarem em mim de olhos fechados.

Às pessoas que Coimbra me trouxe: as minhas canSADAs, Carolina, Catarina e Rafs, pela partilha de desesperos e conquistas nos últimos 5 anos, das tardes de computador à frente, aos momentos de descompressão: "À nOsSa!". À Inês e à Angélica por serem conforto e carinho. Admiro-vos a todas de tantas formas diferentes. Por me mostrarem o que é encontrar o nosso lugar, obrigada.

À Marta, meu pilar incondicional passem os anos que passarem, pela preocupação genuína e ser a minha voz da razão, tem sido uma caminhada bonita contigo.

Àqueles que apesar da distância sei que vibram comigo cada conquista, Maria Inês, Álvaro, Pimenta, obrigada por serem casa.

A todos os outros que, de uma forma ou de outra partilharam comigo os últimos 5 anos, seja em noitadas e tardes pelas bibliotecas de Coimbra, no desanuviar ao fim de uma semana longa ou que tiveram nem que fosse uma palavra de motivação: obrigada por tornarem tudo um pouquinho mais leve.

E por fim, obrigada Coimbra, cidade do coração, por 5 anos numa montanha-russa bonita.

Abstract

Music possesses the ability to evoke a wide spectrum of human emotions, making it a valuable tool for emotional regulation. Therefore, understanding the neural foundations of music-induced emotions is crucial for the development of innovative, personalized neuro-rehabilitative music-based therapy approaches for disorders that lead to an impairment of emotional regulation.

In this study, we employed computational models to identify neural activity patterns associated with perceived emotions while participants listened to diverse musical excerpts. Fifteen participants underwent functional magnetic resonance imaging (fMRI) scans while listening to 96 musical pieces classified by valence and arousal levels (positive or negative), based on a pre-established dimensional model for the categorisation of emotions. The participants also provided their subjective emotional assessments of the music.

We explored different feature selection methods and sets of labels as the classification targets and used multivariate pattern analysis (MVPA) to decode the four emotional quadrants, achieving an average accuracy of 62%±15% in the testing set of the best model. Our findings highlighted the involvement of several neocortical regions (including the auditory cortex, cingulate cortex, somatosensory, motor, and premotor cortices, as well as some visual areas) as important for generating and modulating feeling states.

Keywords: Neuroscience; Music; Emotion; Multivariate Pattern Analysis; Functional Magnetic Resonance Imaging;

Resumo

A música possui a capacidade de evocar um amplo espectro de emoções humanas, tornando-se uma ferramenta valiosa para a regulação emocional. Deste modo, compreender as bases neurais das emoções induzidas pela música é crucial para o desenvolvimento de abordagens inovadoras e personalizadas de terapias neuro-reabilitativas baseadas em música para distúrbios que resultam em comprometimento da regulação emocional.

Neste estudo, utilizamos modelos computacionais para identificar padrões de atividade neural associados a emoções percecionadas enquanto os participantes ouviam diversas peças musicais. Quinze participantes foram submetidos Imagiologia por Ressonância Magnética Funcional (IRMf) enquanto ouviam excertos de 96 músicas classificados consoante os seus níveis de valência e ativação/energia (positivos ou negativos) com base num modelo dimensional preestabelecido para a categorização de emoções. Posteriormente, os participantes forneceram as suas avaliações subjetivas do conteúdo emocional das músicas ouvidas.

Explorando diferentes métodos de seleção de *features* e alvos de classificação, utilizamos a análise multivariada de padrões (AMVP) para classificar os quatro quadrantes emocionais, alcançando uma precisão média de 62%±15% no conjunto de testes do melhor modelo obtido. Os nossos resultados destacaram o envolvimento de várias regiões neocorticais (incluindo os córtices auditivo, cingulado, somatossensorial, motor e pré-motor, bem como algumas áreas visuais) como importantes para a geração e modulação dos estados emocionais.

Palavras-chave: Neurociência; Música; Emoção; Análise Multivariada de Padrões; Imagiologia por Ressonância Magnética Funcional;

Table of Contents

List of Abbreviations

BOLD	Blood-Oxygen-Level- Dependent Signal
CMRO2	Cerebral Metabolic Rate of Oxygen
DLPFC	Dorsolateral Prefrontal Cortex
EEG	Electroencephalography
fMRI	Functional Magnetic Resonance Imaging
fNIRS	Functional Near-Infrared Spectroscopy
FPR	False Positive Rate
FWHM	Full Width at Half Maximum
GEMS	Geneva Emotional Music Scales
GLM	General Linear Model
Hb	Haemoglobin
HRF	Hemodynamic Response Function
IFG	Inferior Frontal Gyrus
ITG	Inferior Temporal Gyrus
LSA	Least-Squares All
LSS	Least Squares Separate
LSU	Least Squares Unitary
MEG	Magnetoencephalography
MER	Music Emotion Recognition
MIR	Music Information Retrieval
ML	Machine Learning
MNI	Montreal Neurological Institute And Hospital
MOG	Middle Occipital Gyrus
MRI	Magnetic Resonance Imaging
MTG	Middle Temporal Gyrus
MVPA	Multivariate Pattern Analysis
NAcc	Nucleus Accumbens
OFC	Orbitofrontal Cortex
PCA	Principal Component Analysis
PET	Positron Emission Tomography
PFC	Prefrontal Cortex
OFC	Orbitofrontal Cortex
PCA	Principal Component Analysis
PET	Positron Emission Tomography
PFC	Prefrontal Cortex
RBF	Gaussian Radial Basis Function
RF	Radiofrequency
ROI	Region Of Interest
RT	Repetition Time
POMS	Profile of Mood States Questionnaire
SFG	Superior Frontal Gyrus
SMA	Supplementary Motor Area
SPG	Superior Parietal Gyrus
STG	Superior Temporal Gyrus
SVM	Support Vector Machines
TPR	True Positive Rate
VMPFC	Ventromedial Prefrontal Cortex
VTA	Ventral Tegmental Area

List of Figures

List of Tables

List of Equations

1

Introduction

In this Chapter, we present the context and motivation for this study in Section 1.1 and the expected goals and contributions in Section 1.2. Lastly, the outline of the document can be found in Section 1.3.

1.1. Context and motivation

Research by anthropologists and ethnomusicologists suggests that music has been present in multiple cultures as a characteristic of the human condition throughout centuries, becoming a fundamental aspect of human nature since ancient times [1].

Music's profound impact extends beyond cultural boundaries, permeating the lives of individuals from all walks of life. From the most knowledgeable music experts to the everyday person, music can remarkably be perceived and enjoyed by all. However, amidst this shared experience, questions arise: How does music elicit specific emotions? Why do the same songs have the power to evoke a wide array of sensations in different individuals, or even within the same person, depending on mood state?

The benefits of music are well-documented across various domains, encompassing physical health, social bonding, cognitive development, and emotional regulation. Engaging in musical activities has been shown to positively affect physical rehabilitation, pain management, stress reduction, immune function, and cognitive skills enhancement [2]. Additionally, it is a powerful tool for emotional regulation, enabling individuals to cope with and alleviate negative feelings such as anxiety, loneliness, and stress while cultivating positive moods such as relaxation or arousal [3].

Given its profound influence on human emotions and well-being, it is essential to delve into the underlying mechanisms by which music facilitates emotional regulation. This

includes a comprehensive understanding of the neural correlates of music-evoked emotions, which will contribute to developing neuro-rehabilitative music-based therapeutic approaches particularly for disorders characterized by impaired emotional regulation [4].

As a multidimensional phenomenon, music exhibits various characteristics, including arousal properties, emotional and valence qualities, and structural features such as tempo, tonality, pitch range, timbre, and rhythmicity. These characteristics give rise to functional and neurochemical effects, mediated by mechanisms such as the brain's reward and arousal systems, impacting pleasure, motivation, stress modulation, and arousal levels [5]. Nevertheless, different frameworks highlight the relevance of additional mechanisms such as evaluative conditioning, contagion, rhythmic entrainment, visual imagery, episodic memory, expectancy, and aesthetic judgment. Collectively, these mechanisms contribute to the subjective emotional experience of music [6], [7].

1.2. Research Goals

The present work aims to explore the brain underpinnings of music-evoked emotions and how they relate to the music's characteristics.

In the last decades, non-invasive neuroimaging techniques, such as functional magnetic resonance (fMRI), have become crucial neuroscience tools as they allow the investigation of brain activity and its underlying functional and structural organisation *in vivo*, enabling a better understanding of neural processes and their relation to various cognitive, affective, and behavioural phenomena.

To this extent, multivariate pattern analysis (MVPA) represents an important tool to identify distributed patterns of activity associated with specific music excerpts, and to analyse how different subjects perceive the emotional content of song clips with the same characteristics.

The main objective is to develop computational models, i.e., classifiers, and evaluate their decoding ability regarding the type of emotional content of musical excerpts. Ultimately, we aim to establish a model of the neural correlates of music-evoked emotion in music.

The significant contributions expected of this research project are to answer the following questions:

- Does the analysis of fMRI patterns during the listening of different auditory stimuli allow us to predict:
    - The type of auditory stimuli (music vs. other stimuli such as white noise)?
    - What is the valence level (positive or negative) of a specific music excerpt?
    - What is the arousal level (positive or negative) of a specific music excerpt?
    - What are, simultaneously, both the valence and arousal levels of a specific music excerpt?

- Does the subjective perception of a participant regarding the emotional content (valence and arousal levels) of a song affect the performance of a computational model attempting to decode the characteristics of the listened stimuli?

- Which brain areas most contribute to our decoding model to identify the specific emotional content of a musical stimulus?

- What is the influence that different approaches to the feature selection (i.e., data and hypothesis driven approaches) have in the decoding ability of the models?

1.3. Outline

This document is composed of seven chapters and is structured as follows. Chapter 2 presents the background on emotions in the brain, fMRI analysis and classification using machine learning (ML), which are key to comprehending the research project's motivational and methodological aspects. The literature's state-of-the-art and most important scientific contributions to this work are summarised in Chapter 3. In Chapter 4, we present the experimental protocol and the methods used for data acquisition, processing, classification, and results analysis. Finally, in Chapters 5 and 6, the results are presented and discussed; in Chapter 7, the conclusions and significance of the research findings are assessed.

2

Background

This Chapter introduces the background concepts necessary to understand methodological details used in the present work to classify music-evoked emotions using fMRI images. Section 2.1 presents an overview of some models that can be used to classify emotions. The fundamentals of fMRI imaging technique are present in Section 2.2, and, finally, in Section 2.3, the background concepts of machine learning and supervised classification are addressed.

2.1. Emotions in the brain

According to Mauss et al. (2005) [8], emotions consist of cognitive, subjective, physiological, and motor changes that arise when an individual consciously or unconsciously determines whether a stimulus has a positive or negative value in a given context. These changes establish a pattern formed by the chemical and neuronal responses that occur in the brain. They are often reflected in a specific behaviour reaction that can be perceived by surrounding people (through body and facial responses) [9].

Pursuing a greater understanding of human emotions has led to different emotional theories or models, each offering unique perspectives on the comprehension of affective experiences. Some of the most commonly used models for conceptualizing emotions are the Russell's circumplex model of affect [10], the Geneva Emotional Music Scales (GEMS)) [11], the Thayer's two-dimension model [12] and the Hevner's affective ring.

A brief description of each model is provided in the next section. We will focus particularly on the Russell's model as it represents a key component of the experimental protocol in this study.

2.1.1. Emotion models

Russell (1980) [10] proposed the circumplex model of affect. According to this model, emotions can be described using an unpleasantness/pleasantness dimension (valence) and a high/low arousal dimension (activation).

Russell asked participants to sort 28 emotion words into categories based on perceived similarity and defined a two-dimensional plane based on positive correlations. The multidimensional scaling analysis revealed the two bipolar, orthogonal dimensions, valence and activation/arousal.

A linear combination of these two dimensions, or different intensities of both valence and arousal, can be used to conceptualise each emotion. For instance, *joy* may be described as the combination of strong activation in the brain systems linked to positive valence or pleasure and moderate activity in the neural systems related to arousal. According to this model, all affective states are observable in these two dimensions, but the levels in each dimension vary.

Similarly to Russell's model of emotions, Thayer [12]also defined emotions in a two-dimension model. The author considers energy and stress as the basic dimensions of emotions. In the energy-stress design, contentment is positioned in low energy/low stress, exuberance in high energy/low stress, anxious/frantic in high energy, high stress, and depression in low energy/high stress correspondingly. The model proposes that mood is a result of the interaction between two neurophysiological systems: the activation system and the inhibition system [13].

Hevner's affective ring [14] resulted from a series of experiments unveiling eight distinct clusters of affective adjectives arranged in a circular manner. Adjectives within each cluster shared similarities, and the meaning of adjacent clusters within the circle varied in a cumulative way, reaching a contrast in the opposite position.

The Geneva Emotional Music Scales (GEMS) [11], were developed to address the complexity not satisfied in simpler models like Russell's circumplex. The authors defined aesthetic emotions to better describe the complexity of the emotional states induced by an

individual's experience while listening to music. Through four interconnected studies [11], participants listed emotion terms during a music listening task. The authors used factor analysis to define a nine-dimensional structure of musically induced emotions. Researchers confirmed its generalizability by replicating this structure with different samples and music genres. The resulting taxonomy of musically induced emotions includes the terms: wonder, nostalgia, transcendence, tenderness, peacefulness, power, joy, sadness, and tension.

Other models that aimed to describe emotional states and their rationale can be found in [15].

2.2. Imaging Techniques

While the categorisation of emotions plays a pivotal role in comprehending music-evoked emotions, it is equally crucial to delve deeper into the neural processes underlying these emotional experiences. In this pursuit, various neuroimaging techniques have been employed to explore the dynamic activation patterns of brain regions associated with emotions during different tasks and stimuli. Techniques such as Functional Magnetic Resonance Imaging (fMRI), Electroencephalography (EEG), Magnetoencephalography (MEG), Positron Emission Tomography (PET), and Functional Near-Infrared Spectroscopy (fNIRS) have each played essential roles in shedding light on emotional neural responses.

In the current work, fMRI was the technique chosen to analyse the interplay between music, emotions, and the brain. Therefore, the forthcoming sections will explore the details of fMRI technique and analysis methodology.

2.2.1. Functional Magnetic Resonance Imaging

Since its discovery in 1990, fMRI has become one of the most powerful techniques for evaluating brain function in clinical and research contexts. FMRI, unlike other imaging methods such as EEG or fNIRS, provides a high compromise between temporal and spatial resolution, with whole-brain coverage.

This technique is based on magnetic resonance imaging (MRI), which creates structural images based on the principle that different tissues' structures present specific magnetic properties.

The critical component of the MRI framework is a strong magnetic field (B0) created by a superconductive magnet. Certain nuclei widely present in biological samples, like hydrogen (protons), have a magnetic moment due to their *spin.* When the protons are placed in the B0 magnetic field, the spin of these elements starts rotating around the B0. If we consider a large sample of protons, let us say a biological tissue (water molecules), the net magnetic moment of a small percentage of hydrogen nuclei (proportional to the water content of the tissue) aligns with the main magnetic field vector B0.

When a radiofrequency (RF) pulse at a specific frequency (dependent on the element and B0 strength) is added, the nuclei resonate and change their orientation. After the RF is turned off, the nuclei return to their original alignment (relaxation or realignment of the net magnetization vector) in a process that involves two main time components: T1 (spin-lattice relaxation time) and T2 (transverse relaxation time) and returns to its previous, resting state causing a signal in the form of a radio wave to be emitted and detectable in a surrounding electrical circuit (by a suitable antenna or coil).

Each tissue type possesses unique characteristics, resulting in distinct signal intensities that depend on the local magnetic properties and interactions.

The images created by MRI can incorporate different types of contrast, such as T1 (enhances the signal of the fatty tissue and suppresses the signal of the water), T2 weighting

(enhances the signal of the water), as well as T2* (time constant of transversal decay when local field inhomogeneities are present) [16].

FMRI extends the information from the static structural images provided by MRI, allowing the characterization of functional changes in the brain. The fundamental basis is to detect neural activity-related changes in blood flow and oxygenation. This detection is done based on the differences in the magnetic properties between arterial and venous blood, which originate the blood-oxygen-level- dependent signal (BOLD) [17], [18] .

When a brain region increases its activity (engaged in a task or behaviour), the additional neural firing results in up-regulated cerebral metabolic rate of oxygen (CMRO2) due to a local increase in energy requirements. Consequently, the oxygen transport molecule, haemoglobin, becomes deoxygenated, and its magnetic properties change [19].

These changes in the haemoglobin properties allow information retrieval by changing the T2* parameter. The extent of these inhomogeneities depends on the physiological state (and, as such, on the composition of the local blood supply). Since the neuronal activity relies on the physiological state, T2* can be considered an indirect indicator of brain activity and is, therefore, the main parameter used in fMRI to create the BOLD contrast [20].

The brain's reaction to a stimulus/behaviour over time, i.e., the temporal response function of the BOLD signal, can be described by an approximation called the hemodynamic response function (HRF). The HRF has been widely studied before, and heterogeneities across the cortex of an individual, between individuals, and across different sensory, motor, and cognitive tasks have been reported.

The BOLD HRF has three phases: the signal intensity increases about two seconds after the stimulus, rises to a peak 6 to 9 seconds later, and then decreases to baseline. It is also common to visualise a post-stimulus undershoot (**Figure 1**).

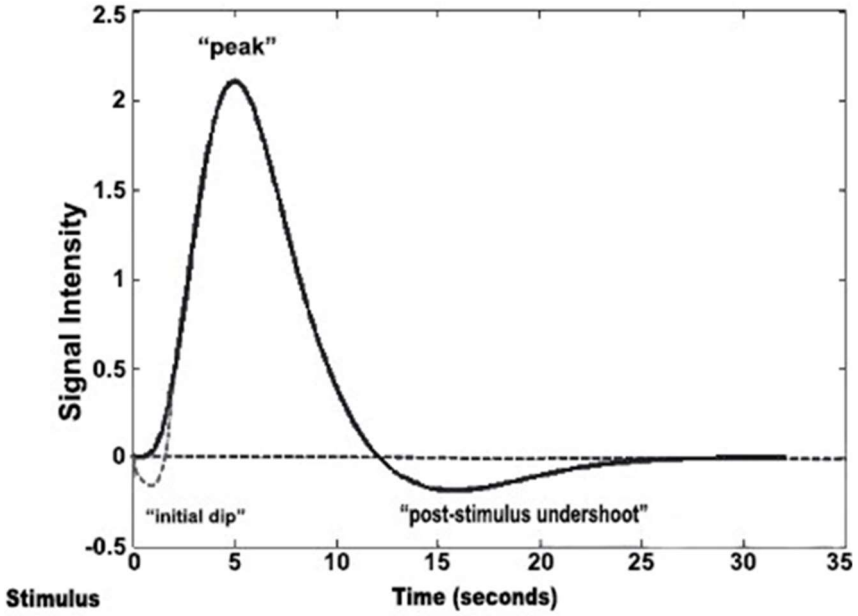

Figure 1*.Hemodynamic response following a short stimulus (adapted from [21])*

2.2.1.1. fMRI images preprocessing

In addition to the neuronal signal previously described, the fMRI signal is also composed of non-neuronal components that include head motion, physiological contributions, tissues that are not of interest, and MRI-induced artefacts. These non-neuronal components limit our ability to interpret the underlying mechanism of brain function and should be addressed. The data quality is also highly dependent on the image acquisition parameters and their influence on the range of intensity values, matrix sizes, and orientations. Therefore, preprocessing steps for fMRI images are required to address these issues.

The most common preprocessing steps are described in **Figure 2** and briefly discussed in this section [22].

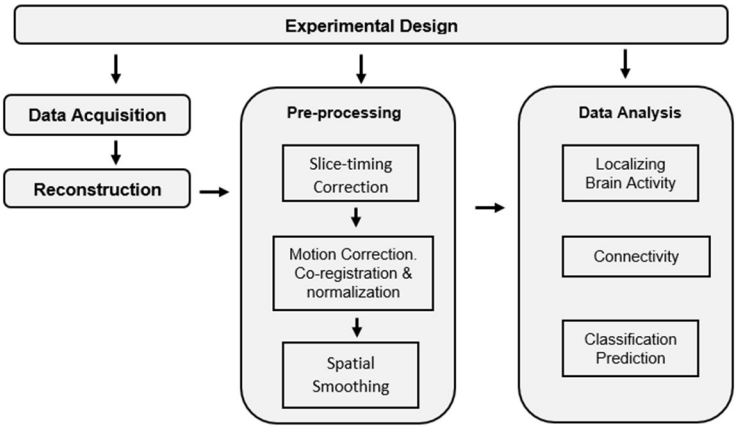

***Figure 2**. The fMRI data processing pipeline illustrates the steps involved in a standard fMRI experiment (adapted from [23])*

Slice-timing correction

One of the principal assumptions while analysing fMRI images is that the whole brain image was acquired in a single shot. However, the complete image is based on a sequentially slice-by-slice reconstruction and, therefore, similar time courses for different slices will be shifted from one another in time.

If this is not corrected, subsequent statistical analyses will present suboptimal results and, therefore, using either interpolation or the Fourier shift theorem, each volume needs to undergo a slice-timing correction to shift each voxel's time course. Interpolation or the Fourier shift theorem can be used to account for variations in acquisition times [23].

Motion Correction

When performing an fMRI experiment, minor displacements occur frequently, even if all the precautions and recommendations to minimize head movements are followed. If this motion is not mitigated (or if the subject is not removed from the study, which can happen if the disruptions are too severe), it can constitute a considerable source of errors and ultimately render the data useless.

Usually, the first or the mean functional image is selected as a target image, and a rigid body transformation considering six parameters (three translation directions and

three rotation types) is used to find the best possible alignment between the target and the other volumes. This matching procedure is carried out by minimising a cost function (such as sums of squared differences) that evaluates how similar the two images are. Lastly, the corresponding transformation, minimising the estimated displacement, is applied to each brain volume [23].

Co-registration

The acquisition of fMRI images is often carried out together with the acquisition of an initial anatomical image (T1 or T2 weighted) that allows a map between what is seen in the lower-resolution functional images and the actual anatomical structures they correspond to, as well as to guide the normalisation of the functional images. This mapping procedure is done by computational algorithms that use either a rigid body (6 parameters) or an affine (12 parameters) transformation together with minimising a cost function [23], [24].

Normalisation

Normalisation aim is to register each subject's images to a standard stereotaxic space defined by a template brain (the Talairach and Montreal Neurological Institute and Hospital (MNI) spaces are the most widely used) to mitigate the anatomical differences between subjects, e.g., brains of different sizes or variations in sulci or gyri. The use of a common standard, combining data across individuals in multi-subject studies allows the comparison of spatial patterns [23], [24].

Spatial Smoothing

Spatial smoothing is standard in the fMRI preprocessing pipeline, replacing the signal at each voxel with a weighted average of that voxel's neighbour's signals. Even though the smoothing of the images may decrease their resolution and blur the image, this procedure provides a cancellation of the noise present in the image and a consequent enhancement of the overall signal, and it may overcome residual anatomical differences that remain and improve inter-subject registration.

Smoothing typically involves the convolution of the images with a Gaussian filter, often described by its specific Full Width at Half Maximum (FWHM) [23], [24].

Temporal Filtering

Due to scanner instability and physiological noise, the signal in fMRI often drifts slowly over time and, as a result, the low-frequency part of the signal has the most power. These drifts significantly diminish the analytical ability of statistical data when they are not taken into consideration and may additionally disprove event-related averaging, which relies on time courses that are assumed to be stationary and have a constant signal intensity.

These variations are eliminated by applying a high-pass filter to each voxel's time course to eliminate variations below a specific frequency cutoff, consequently erasing the impacts of drift [23], [24].

2.2.1.2. fMRI Analysis

After preprocessing, functional data allows us to probe mechanisms, explore temporal and spatial patterns, and ultimately confirm *apriori* hypotheses/theories regarding brain function. There are two main approaches to functional data analysis – univariate and multivariate methods.

2.2.1.2.1. Univariate Analysis

The General Linear Model

When testing hypotheses about spatially constrained, specific effects in neuroimaging data, statistical parametric mapping (i.e., Spatially extended statistical processes) is the most common approach.

In particular, the general linear model (GLM) is an extension of a simple linear regression, and it has been widely used for analysing task-based fMRI images. The method considers an univariate dependent variable, the time course of each voxel, in which a

separate statistical analysis is performed. Therefore, the GLM representation of an fMRI experiment follows a simple linear equation given by:

Equation 1. *General Linear Model equation*

$$Y = X\beta + \varepsilon,$$

where Y corresponds to the fMRI signal from each voxel (1D vector with n time points), X specifies the linear model to be evaluated, with m columns of length n, reflecting each specific factor/regressor hypothesized to influence the dependent variable of the system and the β parameter indicating the amplitude (relative contribution) of each model factor. The final term, ε, corresponds to an error estimate.

The regressors matrix X may include both regressors of interest, also known as task predictors, or nuisance regressors. The task predictors consist of predictions of what the hemodynamic response function (HRF) should look like if a voxel of interest is active as a consequence of a task or stimulus (convolution of the estimated HRF with the design task/stimulus function). Nuisance regressors, account for experimental factors such as head motion or signal drifts that may induce confounds in the analysis. Usually, these confounds are either empirically determined (e.g., retrieved from the preprocessing head motion step) or estimated by modelling a function.

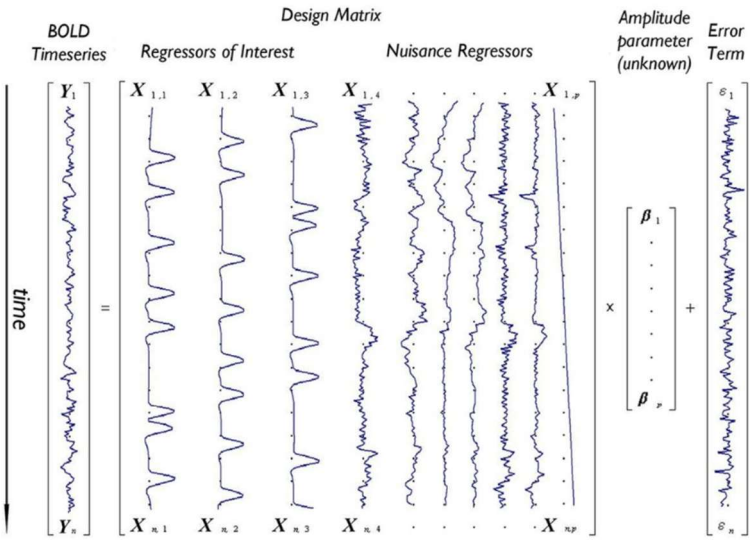

***Figure 3**. Depiction of the General Linear Model (GLM) for a voxel with time-series Y predicted by a design matrix X (including seven nuisance regressors – e.g., six motion parameters and one linear drift). The calculated weighting factors (β 1 – βp) corresponding to each regressor are placed in amplitude vector β while column vector ε contains calculated error terms (ε_i) for the model corresponding to each time point i. (retrieved from* [25]*.)*

Beta-series modulation

The parameters of the GLM are estimated using a cost-function (least-squares error) that minimises the squared error between the time series recorded in each voxel and the predicted time series. [26]

The beta coefficients obtained as an output of the model reflect how much of each voxel's activity can be attributed to the individual stages of a task (i.e., the cue, delay, and probe phases). In fMRI analysis, these coefficients may be arranged in a matrix format and turned into statistical parametric maps of brain activity that are known as beta series (or beta maps).

There are different approaches to create these beta maps that take into consideration the known trial onsets, assumptions about the shape of the BOLD impulse response, and assumptions about noise in the fMRI data [26]:

i) Least-Squares All (LSA)

Each trial of each condition is modelled as a separate regressor in the GLM. This method accounts for variability across different trials of the same type, thus minimising the squared error across all regressors (hence the name least-squares all) [27]. The corresponding design matrix will contain a single regressor for each trial in the run, where each regressor is an impulse function convolved with the hemodynamic response function.

This approach is usually used when there are more brain volumes than trials (i.e., the stimulus onset asynchrony is longer than the TR) because, since every event has its own regressor, in fast event-related designs (e.g., designs where the events occur between 3-6 seconds apart on average) there would be an overlap of regressors [26].

ii) Least Squares Separate (LSS)

Runs a distinct GLM for every trial, modelling each trial as the relevant regressor while combining all other trials into a single nuisance regressor. The design matrices have two regressors: one for the trial of interest and another that simultaneously models all other trials.

When analysing responses to individual trials, the decision between using LSA or LSS should rely on the ratio of trial variability to scan noise: Abdulrahman and Henson (2016) demonstrated that, even in fast designs, when scan noise is higher than trial variability, the LSS model will do better [26], [27].

iii) Least Squares Unitary (LSU)

This is the most typical way of modelling the GLM and does not distinguish between different trials of the same type therefore all trials are collapsed in one single regressor. This method fails in runs with higher inter-trial variability as this variability is relegated to the GLM error but can be used to estimate the mean response for each trial-type [26].

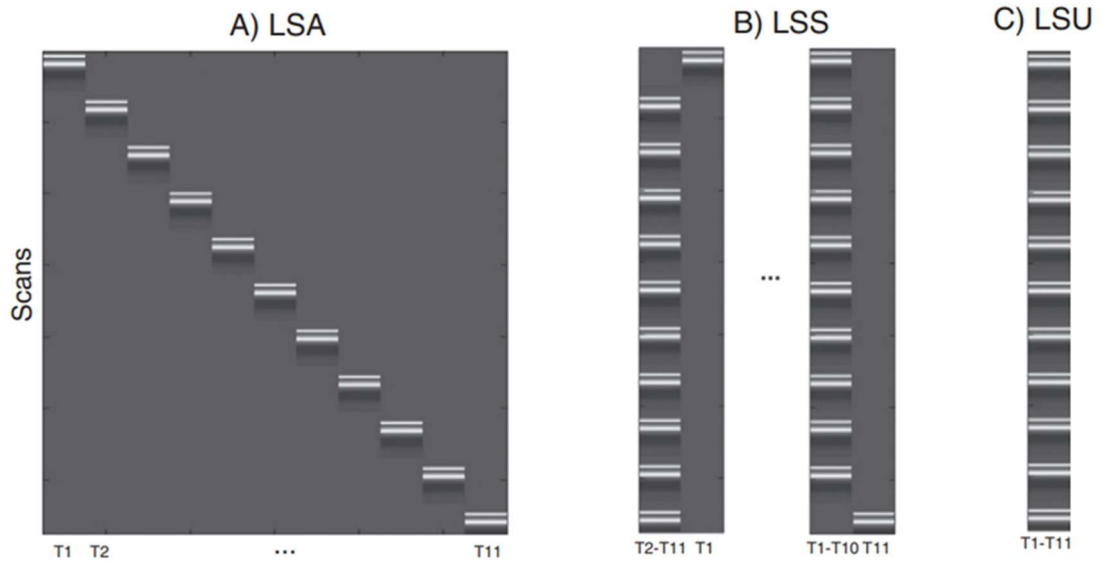

Figure 4. *Design matrices for (A) LSA (Least Squares-All), (B) LSS (Least Squares-Separate) and (C) LSU (Least Squares-Unitary). T(number) = Trial number (Retrieved from [24]).*

2.2.1.2.2. Multivariate Pattern Analysis (MVPA)

GLM's is a powerful, intuitive, and highly flexible tool for analysing brain images and, as an univariate method, it assumes that each voxel is independent. However, brain mechanisms often present multivariate, synchronous activity with nonlinear connections between multiple brain regions. In this sense, multivariate methods are particularly interesting to accommodate fMRI data analysis and may offer advantages compared to univariate approaches [28].

Previous research has established that specific mental states depend on spatially distributed patterns. Univariate methods limit our ability to understand these spatiotemporal patterns. In this sense, two voxels may encode information through their joint activity even though they do not appear to be connected to an experimental variable when examined independently. These dependencies are ignored by univariate analysis. Moreover, the constant increase in the spatial resolution of fMRI provides better access to representational information contained in fine-grained activity patterns in unsmoothed fMRI data [27].

Feature Selection in MVPA

The multivariate analysis aim is to assess multiple variables (also known in the ML literature as features) measured under different circumstances (or classes) and identify how certain combinations of features are related to specific classes. When looking at fMRI data, MVPA usually considers the different experimental conditions as the classes. In the most straightforward scenario, features are single-trial BOLD signal levels in various voxels or estimations of a GLM's parameters (i.e., beta-maps) [27].

MVPA involves searching for highly reproducible spatial patterns of activity that differentiate across experimental conditions. To this end, the definition of spatiotemporal features that the pattern analysis algorithm should use is critical.

In summary, MVPA is a powerful approach to explore the neural underpinnings of emotions in the brain. By transcending the limitations of univariate methods, MVPA enables the capture of distributed brain activity patterns associated with different emotional states. Since the foundation of this technique relies on classification tasks, in the forthcoming sections we will delve into the fundamental concepts of machine learning and classification, and feature selection.

2.3. Classification

2.3.1. Machine Learning

The term machine learning describes the set of techniques and methods that allow machines to improve their performance on a task through experience, both by learning from examples and by identifying patterns in data. ML aims to build models that can automatically detect patterns in data, make predictions or decisions based on those patterns, and continuously improve their performance over time.

Within neuroscience and fMRI analysis context, these models have been used to train classifiers on decoding stimuli, mental states, behaviours, and other variables of interest. The characterisation and interpretation of these models allow for uncovering patterns and relationships that could not be immediately apparent through other methods [29].

Depending on how the learning process was conduced, ML models can have a **supervised** or unsupervised learning. Unsupervised learning aims to find patterns within the data without being explicitly informed on the class of each sample, often through techniques such as compression, dimensionality reduction, and clustering. On the opposite, in supervised learning, the input (or training) data have labels assigned *a priori*. The model aims to understand the relationship between the labels and the data so that when an unknown data sample (feature vector) is given, it can predict its label [29], [30, Ch. 12].

Since the work presented here focuses on a supervised learning classifier, we further detail this model type.

In brief, supervised classification has two main stages, training and testing. To prevent the inflation of classification results, the input data is usually divided into two sets, one for each stage. Data split should consider the amount of available data, leakage (for example, using two samples of a temporal feature close to each other in different sets may ease the classification problem if autocorrelation is high), etc. The split of the data is addressed in the next section.

The training phase corresponds to the learning stage, in which we aim to create an informed decision function based on the training set. Then, the trained model takes the values that various features (independent variables) have in a specific context or example and based on the previously defined decision function, predicts the class to which that example belongs. Overfitting can happen when the created model is too complex and highly adjusted to the training set. This means that the model learns very specific relationships between the training input set of data, and the training labels sometimes related to noise or other irrelevant features, thus performing really well in the training stage, but leading to a deficient capacity of generalisation for new data (i.e., in the testing set). Conversely, if the

model is too simple or does not have enough features, it can fail to capture intrinsic patterns between the data, also leading to poor performance in the testing data. This is called underfitting [31]**.** Therefore, to prevent these phenomena both the choice of the features and the model implemented are crucial factors that one must consider when performing a classification task. Finally, the metric used to determine how well it is performing must also be considered attentively.

2.3.2. Training, testing and validation split

When developing a classification model, the main goal is to perform well in unknown data. To achieve this, the datasets may be divided into two groups: one for training and validation, and the other for testing. The first group contains part of the data, and it is given to the model along with the correct label of each input, being used to train the classifier to learn the patterns underlying each type of label. The validation set is used to keep track of how well the model is doing as it learns, providing an estimate of the model's performance. Finally, to ensure that the model learns these patterns well and its performance is generalisable to new data, the classifier uses the remaining data (i.e., testing set) is used by the classifier to make predictions about what class each input corresponds to. These predictions are then compared to the known labels, and a performance measure gives information about how well the model performed. However, instead of separating the training and validation data into only two groups, a common approach is to do cross-validation by splitting the data into multiple folds and then using each fold as a validation set while the rest are used for training. This is repeated until all folds were considered as a validation set (**Figure 5**).

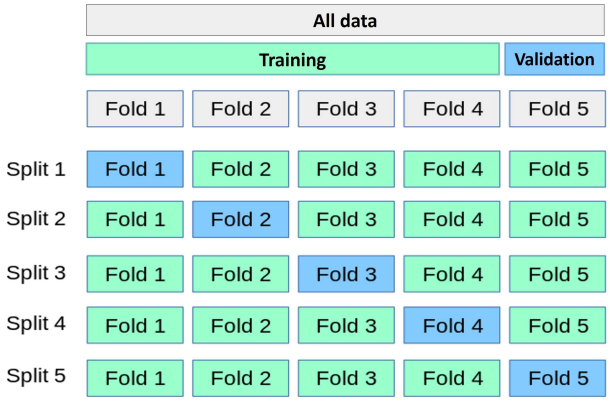

***Figure 5**. Cross-validation scheme (adapted from* [32]*)*

Frequently, the data is first separated in two sets: training and testing sets, and the cross-validation is then performed only in the training set as a way of optimizing the model's hyperparameters. The optimal parameters are then used to retrain the whole training set and the remaining data that was not used to train the model is again used alone to test the final classifier's performance. A common classification pipeline is presented in **Figure 6**.

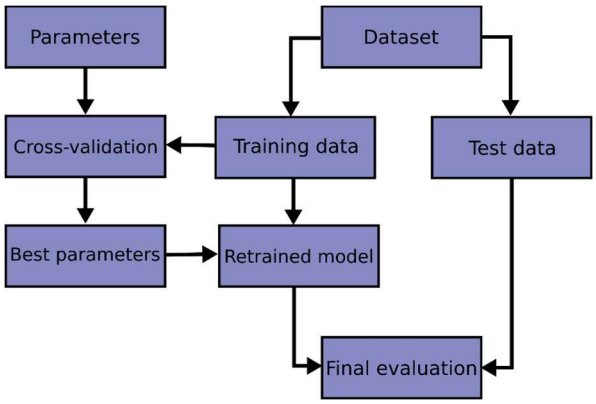

***Figure 6.** Common classification pipeline (Retrieved from* [32]*)*

This method makes a better use of the available data when compared to a two-group separation alone. It is very efficient in preventing and reducing overfitting, reducing bias as well as in providing a more reliable estimate of the performance of the classifier as it is evaluated on multiple times during training[32], [33]

2.3.3. Feature Selection

In ML, when working with high dimensional data, as the number of features or dimensions relative to the number of samples available for analysis increases, the amount of data required to explore and model the space effectively grows exponentially. This is commonly referred to as the curse of dimensionality in ML and can cause problems such as increased computational complexity, sparse data distribution, overfitting, and decreased generalisation performance. It is important to note that these methods should be based on the training set.

Therefore, feature selection (i.e., the reduction of dimensionality of the data, optimising the number of features given as input to the classifier, keeping only a subset of particular interest for the classification problem, is one of the fundamental steps to alleviate these challenges, extract meaningful patterns and insights from the data and obtain good classification results.

Generally, feature selection has three main approaches: filter, wrapper, and embedded-based methods. Filter (or scoring) methods perform statistical tests to assess the relevance of each feature and rank them based on some criteria, such as correlation with the target variable or variance, selecting the top-ranked features to be used in the model. Wrapper methods start with a subset of features and train the model using them. Based on their impact on the classifier more features are either added or removed. Lastly, embedded methods combine characteristics of the previous two [29], [34].

2.3.3.1. Feature selection in fMRI images

The features in fMRI-based classification studies are derived from the BOLD values of each voxel, and therefore the number of available features is always extremely superior to the number of samples. This imbalance increases even more the need for an adequate feature selection. However, in this context, the features are often selected using methods

based on practical experience and observation rather than formal theory. Feature selection is usually done following one of the three approaches:

i) Use of voxel data from the whole brain

Using all available voxels in the feature set maximises the information in the dataset. A potential limitation of this feature set is a dimensionality problem as the number of voxels is highly superior to the number of samples (i.e., time points). The discrepancy between the number of dimensions and samples may lead to an inflated correlation between the activity in many voxels and the labels.

This limitation can be mitigated using different feature engineering techniques. One approach to overcome this problem is preprocessing the data using a principal component analysis (PCA) and setting a feature set with fewer voxels/features. In PCA, the data is linearly transformed into a new coordinate system that uses fewer dimensions than the initial data to describe its variation, thus enhancing the interpretability of data while preserving the maximum amount of information [28].

ii) Use of voxel data from a region of interest defined anatomically or localizer-based

Apriori information regarding brain regions of interest represents a possible alternative to whole-brain analysis. The rationale is to use functional or anatomical masks to reduce the spatial search area, reducing the feature set dimension. This directs the analysis to specific clusters of interest known to be associated with the specific task/behaviour and reduces the computational requirements as the algorithm uses smaller feature sets [26].

When the ROI is defined based on anatomical criteria, a mask is drawn on structural images collected during the MRI session and specific anatomical landmarks, relying only on the expectations about the participation of certain brain areas in a task, hence not needing any functional activity map and providing an unbiased estimate of activity at a given brain area.

On the contrary, functional ROIs include voxels activated by a particular stimulus (the spatial masks are based on the statistical activation map derived from functional data).

Additionally, it is possible to constrain even more the selected voxels so they provide the most valuable amount of information. Therefore, within either a whole brain or a ROI analysis, voxels may be chosen in different manners: based on their activity in at least one condition compared to a control-task baseline, rating each voxel according to the difference in mean activity level between condition and baseline as determined by a t-test (activity-based methods), by scoring them accordingly to how accurately a Gaussian Bayesian classifier can predict the condition of each example in the training set (accuracy method), considering either single voxel's data or data from a voxel and its adjacent neighbours in three dimensions to train the classifier (searchlight). Another method is to use an ANOVA statistical test to identify voxels with significant differences in mean values across conditions. Finally, the stability method selects voxels that consistently show the same activation pattern across the different conditions in the training set, every time these conditions are presented [29].

iii) Searchlight method

The searchlight method considers a small group of voxels centred on a brain voxel, typically a sphere of a given radius, where the centre voxel is moved through the brain so that one pattern analysis is performed for each possible location. Contrary to whole-brain and region-of-interest analyses, which result in a single value, the searchlight analysis gives one value for each centre voxel location. A classification result can be attributed to each centre voxel to form a statistical map of local multivariate effects. For this reason, searchlight analysis is also referred to as mass-multivariate analysis [27].

Creation of feature vectors

We have previously described how features can be created based on fMRI data, either the BOLD signal values or GLM parameter estimates. These data are then reorganized into

feature vector through the concatenation of multiple voxels into a voxel activation vector (x= x1+ x2+...+xp, with p voxels), each feature vector associated with specific experimental condition (i.e., target or label). It is possible to interpret the elements of these vectors x as coordinates in a high-dimensional voxel activation space. Clusters of data points arise in this space due to the similarity of multivariate responses within a condition and the contrast between conditions, which the pattern analysis algorithm can distinguish. In **Figure 7** it is shown an example of the definition of voxel activation space with only two dimensions.

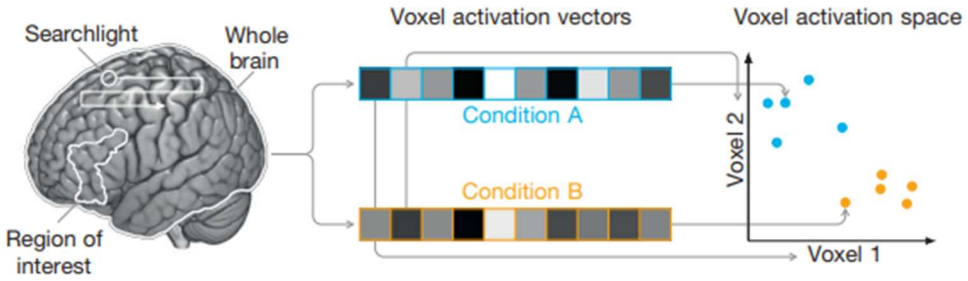

Figure 7. *Definition of voxel activation space with two dimensions. (Retrieved from* [35]*)*

2.3.4. Support vector machines (SVM)

Support vector machines (SVM) is a popular supervised classification method that maps the input data into a high-dimensional feature space and then finds the hyperplane that separates the data points into different classes. The name of this method is given by the data points that lie closest to the hyperplane (or decision surface), called support vectors, which are the most difficult to classify and have a direct bearing on the optimum location of the decision surface.

The optimisation problem is the maximisation of the distance between the closest data points of different classes and minimisation of the classification errors.

When there are only two classes and the data is linearly separable, the hyperplane can be defined by:

Equation 2. *Hyperplane Equation in a linearly separable problem*

$$w^T X + b = 0\,,$$

where w is the weight vector, normal to the hyperplane thus determining its orientation, b is a bias term (determines the position of the hyperplane relative to the origin by shifting it along w by a distance of b units) and $X = [x_1, x_2, x_3 \dots x_n]$ is the training dataset with n points that may belong to each class y_i ($y_i = \{-1,1\}$). [36] The classification of the data points is made according to:

$$y_i = 1 \; if \; w^T X - b \geq 1$$

$$y_i = -1 \; if \; w^T X - b \leq -1$$

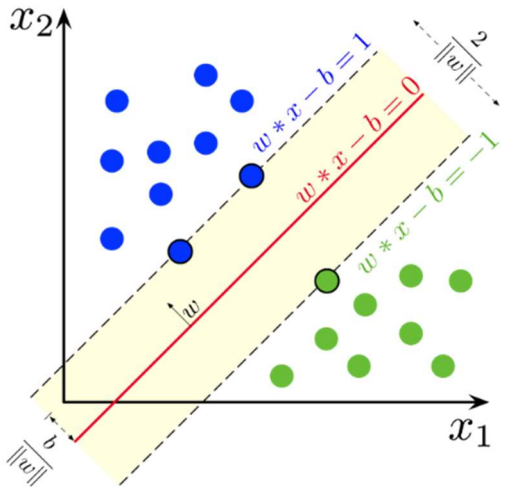

Figure 8*. Maximum-margin hyperplane and margins for a linearly-separable SVM trained with samples from two classes.(Retrieved from [37])*

In more complex contexts, usually there are more than two labels to attribute to the data. In these situations, the two main multi-class classification methods are:

- **One-vs-all:** N binary classifier models are generated for each of the N class labels of the dataset. Each classifier is trained considering all the examples of

the i^{th} class with a positive label and all the other with a negative label. In the prediction phase, the data is considered as input for all generated classifiers and the decision function of each one is evaluated. The classifier that provides the highest output value determines the final predicted class [36], [38].

- **One-vs-One:** N(N-1)/2 binary classifiers are generated for each pair of N classes. The input data is then classified by all models that predict one label. The labels predicted by each classifier are counted and the final output is given by the class with the highest count [36], [38].

Additionally, in most real-life situations, the input data is not linearly separable, leading to a classification problem that needs to address nonlinearity. One of the approaches is to allow for some points to be misclassified by introducing a "slack variable" that allows some training examples to be on the wrong side of the hyperplane and tries to find a hyperplane that separates the data with the smallest possible number of errors. This is called the soft margin approach and is helpful when handling noisy or overlapping data.

This “slack variable” is often referred to as the C parameter and determines the penalty for misclassifications. Therefore, a smaller value of C allows for a larger number of misclassifications and a wider margin, producing a simpler model but with higher generalisation ability. In comparison, a larger value of C penalises misclassifications more heavily, leading to a narrower margin and a more complex model.

The choice of the C parameter depends on the specific dataset and problem at hand and should be done when setting the model’s hyperparameters. It is typically determined using techniques like cross-validation or grid search, where different values of C are evaluated, and the one that results in the best performance on a validation set is selected [39].

SVM also implements the kernel trick: a kernel function transforms the input features into a higher-dimensional space, the feature space, where the data becomes linearly separable (**Figure 9**). The data are still nonlinear in the input space while a linear SVC

can be created in the feature space to separate them. A map $\phi: \mathcal{R}^n \mapsto \mathcal{H}$ is chosen, where the dimensionality of $\mathcal{H}$ is greater than n.

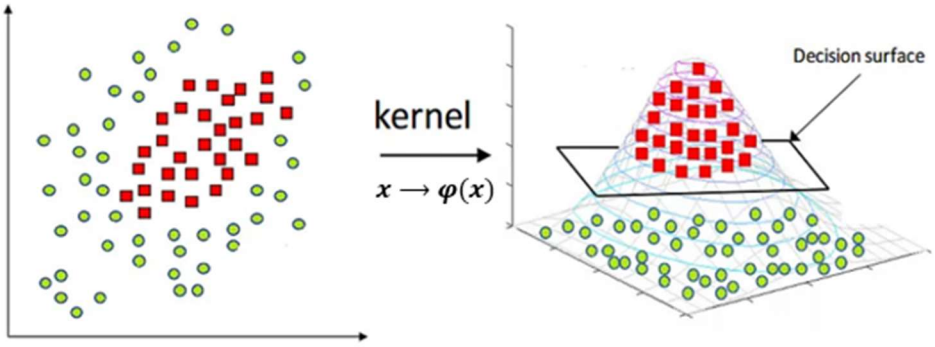

***Figure 9**. Kernel transformation of non-linearly separable data.(Adapted from* [40]*)*

This is done using kernel functions with the form $\kappa(x_i, x_j) = \phi(x_i) \cdot \phi(x_j)$ to transform the data.

The most used kernel functions are summarized in **Table 1** [39].

***Table 1**. Different Kernel Functions Often Used for Nonlinear Data Classifications Using a SVC*

Kernel Function	**Type of Classifier**
$\boldsymbol{\kappa(x_i, x_j) = (x_i^T x_j)^{\rho}}$	Linear
$\boldsymbol{\kappa(x_i, x_j) = (x_i^T x_j + 1)^{\rho}}$	Complete polynomial of degree ρ
$\boldsymbol{\kappa(x_i, x_j) = \tanh(\gamma\, x_i^T x_j + \mu)}$	Multilayer Perceptron
$\boldsymbol{\kappa(x_i, x_j) = e^{-\frac{\|x_i - x_j\|^2}{2\sigma^2}}}$	Gaussian Radial Basis Function (RBF)
$\boldsymbol{\kappa(x_i, x_j) = \tanh(\alpha\,(x_i \cdot x_i) + \vartheta)}$	Sigmoid

2.3.5. Other supervised classifiers

Besides SVMs, there is a large number of other commonly used supervised classifiers.

For example, K-Nearest Neighbours assigns a new data point to the class most commonly found among its K nearest neighbours, based on a chosen distance metric [41], [42].

Additionally, decision trees learn a hierarchy of informative if/else tests that form a binary tree, allowing for decisions to be made based on the most valuable features at each step, determined by entropy and impurity measurements, resulting in an interpretable and effective algorithm for classification and regression tasks [33], [41].

A Naïve Bayes classifier looks at each feature individually and estimates a class membership probability, always assuming naively that the values of the different features do not affect each other. It then uses those probabilities to predict to which class a new input is most likely to belong to [33], [43].

Neural networks are complex interconnected models inspired by the structure and function of the human brain's neurons, capable of performing complex tasks by processing input data through weighted connections, activation functions, and multiple layers [33], [41].

More detailed information about all these classifiers can be consulted in the mentioned references.

2.3.6. Performance Metrics

Performance metrics play a crucial role in evaluating the effectiveness of ML models by providing quantitative measures to assess their performance. These metrics are essential to compare different models or algorithms, understand how well a model performs, and generalise on unseen data.

The selection of appropriate evaluation metrics depends on various factors, such as the problem domain, the data, and the desired outcome.

Some commonly used performance metrics include:

Accuracy

The accuracy of a classification algorithm measures the number of correct predictions made as a ratio of all predictions made.

Considering a set with a total of samples to be classified of $n_{samples}$, $\hat{y}_i$ the predicted value of the i-th sample and y_i the corresponding true value, then the accuracy is given by:

***Equation 3**. Accuracy Calculation*

$$accuracy(y, \hat{y}) = \frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}-1} (\hat{y}_i = y_i)$$

Precision

In a binary classification task, precision is the ability of the classifier not to label as positive a sample that is negative. Considering TP as the true positives, FP as the false positives, we have:

***Equation 4**. Precision calculation for binary classification*

$$precision = \frac{TP}{TP + FP}$$

Recall

Recall, sensitivity, or true positive rate (TPR) is the ability, considering a binary problem, of the classifier to find all the positive samples. Considering FN as the false negatives, recall can be calculated with:

Equation 5. *Recall calculation for binary classification*

$$Recall = \frac{TP}{TP + FN}$$

False Positive Rate

The false positive rate (FPR) is a metric that measures a classifier's ability to identify negative samples correctly within the set of all the samples labelled as negative. It represents the ratio of negative instances erroneously classified as positive and the total number of actual negative events. Considering TN as the true negatives and FP as the false positives, we have:

Equation 6. *FPR calculation for binary classification*

$$FPR = \frac{FP}{TN + FP}$$

F1-score

The F1-score metric is the harmonic mean of precision and recall, presenting a good balance between both measures.

A high F1-score symbolizes a high precision as well as high recall. As it is a measure more sensitive to data distribution, it is often use on imbalanced classification problems.

Equation 7. *F1-Score calculation*

$$F_1 = \frac{2}{\frac{1}{precision} + \frac{1}{recall}}$$

However, the precision, recall and F1-score formulas only work for binary classifiers. When performing multi-class classification, each measure is calculated per class in a one-

vs-rest manner, i.e., each class's success is rated separately, as if there are distinct classifiers for each class.

There are then two main methods used to access the value of the overall classifier:

i) Macro-Averaging: division of the measure value of each individual class by the number of total classes, c

***Equation 8**. Calculation of performance measures in multiclass problems with macro-averaging*

$$Measure_{macro-avera} = \frac{\sum_{i=1}^{c} measure(i)}{c}$$

ii) Weighted-Averaging

The weighted-averaged score is calculated by taking the mean of all per-class scores while considering the proportion (weight) of the number of occurrences (n_i) of each individual class, c, relative to the total number of predictions, N.

This is usually the go to evaluation method when the classes are not evenly distributed.

***Equation 9**. Calculation of performance measures in multiclass problems with weighted averaging*

$$Measure_{weighted-ave} = \sum_{i=1}^{c} \frac{n_i}{N} measure(i)$$

3

State of the art

Music is often considered an abstract pleasure with aesthetic value and, therefore, the benefits that come from listening to it and how these emerge may not always be as obvious or easy to comprehend as the ones provided by food or socialising, well known by their role in adaptive evolution. As already mentioned, understanding these mechanisms offers great value for both research and therapeutic endings, having become a popular area of investigation among scientists.

In the following sections, a review will be presented on how and where these effects emerge and the possible benefits of emotional regulation through music.

In Section 3.1 the mechanisms through which music elicits emotions are explored. Research on the brain areas involved in the emotional response to music is reviewed in Section 3.2 and the relationship between specific musical features and emotions is explored in Section 3.3. Finally, in Section 3.4 we examine how music can be used as a therapeutical tool.

3.1. How does music elicit emotions?

Understanding how music elicits emotions combines i. exploring specific emotions linked to musical pieces and ii. the underlying mechanisms through which they are evoked.

We here highlight the BRECVEMA model, initially proposed by Juslin and Västfjäl in 2008 [44], as it includes six interconnected mechanisms through which music may induce emotions: brain stem reflexes, evaluative conditioning, emotional contagion, visual imagery, episodic memory, and musical expectancy (**Figure 10**). Later, in 2013, Juslin [6] revised this

theory and included two additional dimensions: Rhythmic Entrainment and aesthetic judgment.

The **brain stem** (**B**) is an evolutionarily ancient brain structure responsible for various functions, including auditory perception, attention, emotional arousal, heart rate, breathing, and movement. It plays a role in arousal-mediated mechanisms triggered by basic acoustic events like sudden loudness or fast rhythms. These events cause physiological arousal, leading to changes in heart rate, blood pressure, and skin conductance due to the release of stress hormones. This reaction can be interpreted as a response rooted in primal instincts relevant for survival [45].

Evaluative conditioning (**E**) is described by the authors as a type of classical conditioning that involves the repeated association of an initially neutral stimulus (i.e., music) with an affectively valenced stimulus. The initially neutral music acquires the ability to evoke the same affective state as the valenced stimuli in the perceiver. It seems to depend on unconscious, unintentional and effortless processes that involve subcortical brain regions (such as the amygdala) and the cerebellum [45].

Emotional contagion (C) refers to the phenomenon where music evokes an affective reaction that mirrors its own expressive character, either through peripheral feedback from muscles or a more direct activation of the relevant emotional representations in the brain - it elicits an emotional echo in the listener. This process involves the mirror neuron system, located primarily in the premotor cortex, which is thought to simulate an emotional state in the listener by linking perceptual and behavioural representations of a stimulus during the perception of emotionally arousing music [46] and is also influenced by individual characteristics of the listener [47], [48].

Episodic memories (E) are similar to emotional conditioning in the sense of both evoking emotions associated with a certain moment or situation but, in episodic memories, there is a conscious recall of a past event in time, that preserves the contextual information. Next, the effect of **visual imagery (V)** is explained as the process through which a listener experiences a feeling because of visual imagery while listening to music. The interplay between the music and the images closely influences the emotions experienced and is very

common in a music-therapy context. There is a possible connection with episodic memories, but these two mechanisms should be distinguished as emotions may emerge from a visual imagery of a thing that was never experienced.

Musical expectancy (M) refers to the anticipation of upcoming musical events, such as a particular melody or rhythm, and the subsequent satisfaction or surprise when those events are met or not met, involving the reward brain network to arise to feelings of anxiety /frustration and surprise. [49] It is essential to differentiate between the surprise elicited by brainstem mechanisms, which represent a reflexive response to unexpected and potentially threatening stimuli, and the non-fulfilment of expectancies, which involves a cognitive process of predicting subsequent musical events based on previous experiences and knowledge of musical structure [6], [44].

The two additional mechanisms proposed by the authors (2013) [6] are **aesthetic judgment** and **rhythmic entrainment**. The first one is assumed to rely more on higher cognitive functions, domain-relevant knowledge, and a fluid, individualised process that may change across time and context (involves the subjective evaluation and perception of a musical piece's beauty or artistic merit). Rhythmic entrainment is a mechanism through which body movements synchronise with the beat of the music, allowing for individuals to move along with the music. It involves neural processes that allow perceiving and anticipating rhythmic patterns in music and physically engaging with it.

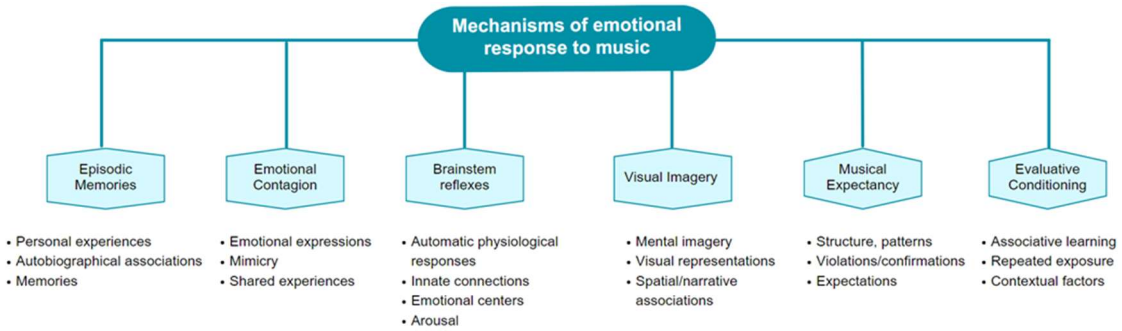

Figure 10. *Six main mechanisms through which music is able to elicit emotions.*

3.2. The brain while listening to music

There are many brain regions working together to shape our emotional experiences. Understanding the interplay between these regions provides valuable insights into the complex nature of music's emotional impact on our brains.

In particular, the brain's reward system has been frequently pointed out as comprising a crucial group of regions and structures involved in the processing of music evoked emotions.

In fact, a metanalysis conducted in 2020 by Stefan Koelsch [4] on emotions elicited by music identified clusters along the whole reward network as relevant to these processes.

3.2.1. Reward Network

The reward system of the brain is a complex network of brain regions and neural pathways, responsible for processing rewarding or pleasurable experiences and reinforcing behaviours associated with positive outcomes (**Figure 11**).

This network relies primarily on the dopaminergic neurons located in the midbrain, particularly the ventral tegmental area (VTA), and their projections to key areas, operating through two main pathways: the mesolimbic and the mesocortical pathways [50].

The mesolimbic dopamine pathway starts with the production and release of dopamine in the VTA and projects mainly to the nucleus accumbens (NAcc) in the ventral striatum, an area associated with motivation, but also into the amygdala and lateral hypothalamus.

The mesocortical pathway connects the VTA to the prefrontal cortex (PFC), also including the orbitofrontal cortex (OFC), a key area involved in cognitive processes, such as decision making and memory [51], [52].

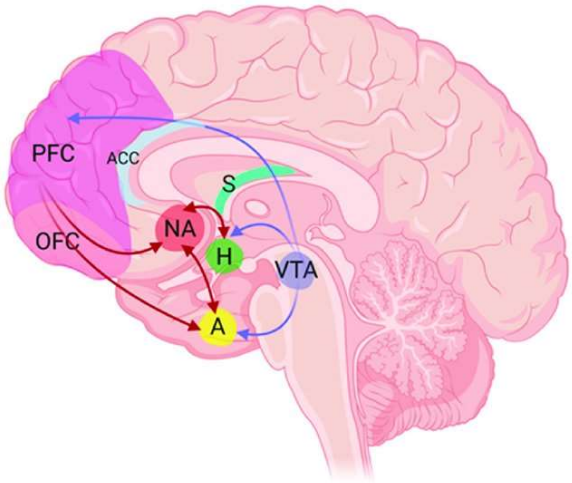

***Figure 11**. Reward pathways. VTA, ventral tegmental area; NA, nucleus accumbens; H, hypothalamus; PFC, prefrontal cortex; OFC, orbitofrontal cortex; A, Amygdala; ACC, anterior cingulate cortex; S, striatum. (Retrieved from [50])*

The hippocampus

The hippocampus is commonly known for its involvement in cognitive functions like memory, learning, and spatial orientation [53]. However, recent research has shed light on its contribution to emotional processing, extending to the realm of music. In fact, the Koelsch (2020) meta-analysis revealed significant clusters in the anterior hippocampus, with a prominent right laterization, relevant in music-evoked emotions.

Notably, studies investigating the effects of pleasant and unpleasant music in the brain have shown that the hippocampus is activated during the perception of unpleasant songs and deactivated during the listening of pleasant songs [25], [54]. This activation pattern suggests that the hippocampus plays indeed a crucial role in processing emotional content associated with music, particularly with negative valence.

Consequently, these findings strongly suggest that the activation of the hippocampus in response to music stimuli is not solely limited to cognitive processes but also plays a significant role in the processing of emotions.

The ventral striatum and the nucleus accumbens

The NAcc is a key component of the brain's reward system, located in the ventral striatum and plays a vital role in motivation, pleasure, reinforcement, and addiction [55].

As a matter of fact, the NAcc was pointed in various studies as a fundamental structure related, in particular, to the rewarding nature of music ([56], [57], [58]). It has been reported that it is highly activated in response to pleasant music, a behaviour connected with an increase of the dopamine availability in the ventral striatum and that is not seen in people with musical anhedonia (i.e., the inability to experience pleasure from normally pleasant stimuli even if there is still an ability to perceive and understand music on a cognitive level) [58]. Additionally, the NAcc is also activated during musical prediction errors or during some unexpected moments in a song, highlighting a role in the processing of unexpected and novel music stimuli [56].

The amygdala

The amygdala is a major component of the limbic system, considered the integrative centre for emotions, emotional behaviour, and motivation of the brain [55].

Results show an important role of the amygdala in processing emotions elicited by music, evidenced by the large clusters in this area found in the meta-analysis conducted by Koelsch [4]. These clusters were present despite the type of emotions evoked, i.e., either positive or negative. In particular, similarly to what was found in the hippocampus, a study revealed a strong deactivation of the amygdala during the listening of pleasant songs, contrasting with a hyper-activation during unpleasant musical stimuli [54].

3.2.2. The cerebellum

The cerebellum is usually associated to motor behaviour and the control of balance, but recent studies have suggested its involvement in cognitive and emotional processing,

including music processing [59], as well as in processes related with the reward systems of the brain [60].

The cerebellum is involved in processes related to pitch discrimination ([61], [62]), rhythm and beat interval discrimination, as well as in discriminating time intervals in general and in music perception.

Additionally, impairments in the cerebellum have been used to probe its role in emotion processing, in particular with musical stimuli [63], [64]. These studies revealed a correlation between the damages in this structure and a poor ability to recognise the emotional content of songs thus demonstrating its relevance.

Finally, research on the neural correlates of music-evoked emotions has also identified clusters in the cerebellum relevant for the discrimination of emotions such as fear and joy during the listening of musical stimuli music [65], [66].

3.2.3. Decoding Studies

Despite the undeniable importance of structures within the limbic and paralimbic system and, in particular, within the brain's reward network in the processing of music-evoked emotions, the studies that identify the role of these structures are primarily univariate, using GLM contrasts. However, when using multivariate pattern analysis to decode the emotional content of musical *stimuli,* the results frequently do not include the presence of subcortical structures.

In this section, we present the results of some recent studies that performed MVPA with different methods of classification and feature selection and were able to decode the emotional content of music clips successfully.

Neural decoding of emotions based on music listening tasks

Koelsch, S. et.al (2021) [67] recorded the BOLD signal of participants listening to music clips to evoke feelings from two different categories: joy and fear. The joy *stimuli* were excerpts taken from CD-recorded pieces from various styles (soul, jazz, Irish jigs, classical, South American, and Balkan music) and the fear *stimuli* were excerpts from soundtracks of

suspense movies and video games. The researchers then used the searchlight method (a spherical searchlight with a radius of 3 voxels) with a linear SVM classifier to analyse the brain activation patterns and decode what type of musical stimuli each participant was listening to in a specific moment. They found that the decoding accuracy was higher for joy than fear clips and for feelings evoked by musical features related to rhythm and harmony than those related to melody and timbre. Additionally, a positive correlation between the subjective ratings of the stimuli and accuracy, and between decoding accuracy and musical training was also revealed.

The researchers identified voxels with significant above-chance information for the classification task in multiple neocortical regions: auditory cortex, primary and secondary somatosensory cortex, premotor cortex, frontal operculum. Some significant clusters in the granular insula and cingulate cortex were also found.

A similar study conducted by Putkinen, V. et al (2021) [68] in a large sample of participants (n=102) aimed to explore music evoked emotions using musical clips associated with different emotions: sadness, happiness, fear and tenderness. The music labels were attributed to each excerpt based on their ability to reliably induce the target emotions, as determined by behavioural ratings from participants in a previous study [69]. Additionally, to map neural circuits governing non-musical emotions, the participants also viewed film clips with positive, negative, and neutral emotional content (the emotional ratings were obtained by a separate sample of subjects that viewed the clips and rated the intensity of positive and negative emotions observed).

The MVPA was performed using a SVM with RBF kernel classifier both in whole-brain data and within a subset of the regions of interest (ROIs) where emotion classification has been successful in previous studies (amygdala, hippocampus, thalamus, anterior and posterior cingulate, SMA, precentral and postcentral gyri, precuneus, frontal pole, auditory cortex).

In the whole-brain MVPA, each musical emotion was classified significantly above chance level. In particular, the classification accuracy for happiness and tenderness was slightly lower than for fear and sadness. In the ROI-level MVPA, the classification accuracy

for all four emotions was above chance level for the auditory cortex and reached significance in the precentral gyrus.

Furthermore, in a GLM analysis, the emotions evoked by music were found to consistently engage a network of auditory cortical areas and regions supporting motor control, somatosensorial, and interoceptive processing (pre- and postcentral gyri, SMA, cerebellum, ACC, insula and precuneus). However, contrary to what was verified for non-musical emotions, music did not strongly activate limbic and medial prefrontal regions. This important result suggests that the neural circuits governing non-musical emotions are different from those governing music-induced emotions.

The authors of these studies have identified key regions for the decoding task in the auditory cortex (Heschl's gyrus, superior temporal gyrus (STG)), in the middle temporal gyrus (MTG), superior temporal sulcus, *planum polare, planum temporale*), in regions associated with motor control (in particular the precentral gyrus: the site of the primary motor cortex, and the supplementary motor area) and in somatosensory areas (postcentral gyrus, which is the location of the primary somatosensory cortex. The insula, in particular the posterior insula, which is associated with interoceptive functions was also frequently pointed as containing information for the decoding of the emotional content of different auditory *stimuli* [67], [68], [70]–[72].

Brain regions encoding musical features

In addition to the decoding of music-evoked emotions, researchers have applied similar multivariate analysis techniques to investigate how specific musical features are represented in the brain.

Michael Casey (2022) [79] used spherical searchlight regression analysis to predict brain activity patterns in response to features like melody and harmony across a wide array of cortical areas. The prediction-accuracy maps indicated significant clusters of brain activity in regions spanning the temporal, frontal, parietal, and occipital lobes, as well as the parahippocampal gyrus and cerebellum. Moreover, the study identified specialised regions

responsible for encoding tonal music details in terms of relative pitch representations. This research sheds light on how the brain's architecture processes and represents various fundamental musical features.

3.3. Music information retrieval

Music Information Retrieval (MIR) is a multidisciplinary research field that focuses on the development of information extraction tools from music for multiple purposes including automated labelling, automated annotation, estimation of properties such as rhythm, tempo, etc. Ultimately, this research field aims to make music more accessible, easier to create and describe and analysable.

Music Emotion Recognition

Music emotion recognition (MER) is a subfield of MIR referring to the process of extracting and analysing music features, defining the relationship between music features and a specific emotion space, thus allowing the recognition of the emotional content that a music expresses [73]. The emotion space itself may be defined according to different proposed models such as Hevner's affective ring, Russell's circumplex model of affect [10], the GEMSs [11] or Thayer's two-dimension model [12]. Several audio features have been widely studied and used in MER applications (**Table 2**)

Table 2. *Musical features relevant to MER**

Features	Examples
Timing	Tempo, tempo, variation, duration, contrast.
Dynamics	Overall level, crescendo/decrescendo, accents.
Articulation	Overall (staccato, legato), variability.
Timbre	Spectral richness, harmonic richness.
Pitch	High or low.
Interval	Small or large.
Melody	Range (small or large), direction (up or down).
Tonality	Chromatic-atonal, key-oriented.
Rhythm	Regular, irregular, smooth, firm, flowing, rough.
Mode	Major or minor.

Loudness	High or low.
Musical form	Complexity, repetition, disruption.
Vibrato	Extent, range, speed.

**(adapted from* [74]*))*

Depending on the authors, musical attributes may be divided into four to eight (rhythm, dynamics, expressive techniques, melody, harmony, tone colour, musical texture and musical form) different categories to facilitate the identification of where features related to emotion belong as well as to identify which categories may lack computational models to extract features relevant to emotion from music. A detailed table with multiple examples and descriptions of features that from each of these categories can be consulted in [74]).

Panda et. al (2020) [74] proposed a novel set of emotionally relevant features. To evaluate this feature set, the authors created a dataset including 900 song entries, tagged with emotion labels given by the website from where the songs were retrieved. These tags were posteriorly mapped into Russell's valence and arousal quadrants using Warriners's adjectives list [75] and, finally, the music clips were manually annotated by different subjects in terms of Russel's quadrants to validate the mapping between the labels given by the website.

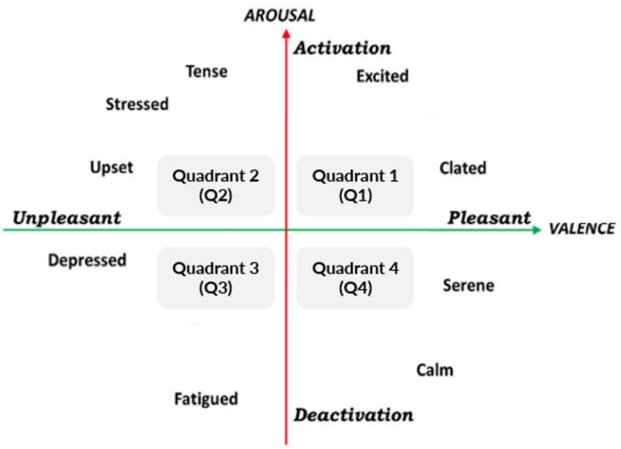

Figure 12. *Graphical representation of the circumplex model of affect (adapted from [76])*

The authors used multiple musical features to decode the quadrant to which a music excerpt belonged to. The classification results from Q1 and Q2 excerpts presented higher accuracies compared to excerpts from Q3 and Q4. Considering the labels proposed by the authors, the results suggest that music clips with higher arousal are easier to differentiate. Conversely, the lowest performance in Q3 and Q4 reflect a higher ambiguity for valence in songs with low arousal. Furthermore, certain compositions within these quadrants exhibit similar musical features that correspond to divergent emotional elements. For instance, some songs may possess a cheerful melody or accompaniment coupled with melancholic vocals or lyrics.

The best classification results were obtained using 29 novel features and 71 baseline features. The authors provide a detailed description of these features and algorithms to compute them.

3.4. Music as a therapy approach

The ability to use music to evoke specific emotions makes it an extremely powerful therapeutic tool.

It has been demonstrated that, in clinical contexts, patients receiving physical and psychologically challenging treatments, may benefit if they receive the treatment while listening to music [77].

If these improvements in well-being can be implemented in a controlled way, individually adjusting the music interface to generate specific emotions, we will ultimately optimize the control loop and achieve best possible outcomes.

Currently, neurofeedback is becoming a very popular method to self-regulate brain function by measuring the brain activity of an individual with common neuroimaging tools such as such fMRI, fNIRS, EEG , etc, and then presenting it to the subject (feedback) [78]. The main goal is to volitionally or non-volitionally control either the activity in a particular brain region or the functional connections between various brain regions.

Numerous studies have consistently reported the activation of brain regions associated with emotion-evoking mechanisms, particularly those involved in reward and valence pathways [79]. Building upon this knowledge, Direito et al. (2021) [80] conducted a neurofeedback experiment aiming to modulate the activity of a specific brain region and investigate the effects of explicit positive and negative valence feedback stimuli on the reward and saliency networks. The findings revealed that positive feedback resulted in increased activity in the ventral striatum, while negative feedback led to increased activity in the anterior insula, two regions previously found to be activated by music in other studies [54], [67].

These observations are particularly significant considering the well-established links between these brain regions and psychiatric disorders such as addiction, depression, and anxiety, as well as the demonstrated ability of music to engage these regions. Therefore, these findings shed light on the therapeutic potential of using neurofeedback interventions targeting the reward and saliency networks, especially in the context of leveraging music's influence on these systems.

In fact, the feasibility of using music in combination with real-time fMRI neurofeedback to up-regulate emotions have already been described by Lorenzetti et al. (2018) [81]. The authors used a combination of music and real-time fMRI neurofeedback to up-regulate tenderness and anguish, which engaged specific brain regions previously implicated in positive affiliative emotions (such as the septo-hypothalamic region, medial frontal cortex, temporal pole, and precuneus) and negative affect (the amygdala, dorsolateral prefrontal cortex (DLPFC), and others), respectively.

These collective findings highlight the promising potential of utilizing neurofeedback techniques, particularly when combined with music, as a therapeutic approach for modulating emotions. Further exploration of this therapeutic field holds promise for addressing various psychiatric conditions associated with emotional dysregulation.

4

Methods

4.1. Participants

Fifteen individuals (9 females, 6 male; age range 22–41 years, M=31.7, SD=6.27) took part in the experiment. All participants gave written informed consent. The study was conducted in accordance with 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.

All the volunteers filled a "Profile of Mood States" (POMS) questionnaire prior to the tasks and reported to have normal hearing, without permanent or current temporary impairments and with no known history of neurological disorders.

4.2. Data acquisition

MR acquisition was performed in a 3 T Siemens Magnetom Prisma scanner with a 20-channel head coil at the Institute of Nuclear Sciences Applied to Health (ICNAS), Coimbra. First, a high-resolution (1 × 1 × 1 mm) T1-weighted anatomical reference image was acquired from each participant using MPRAGE sequence [82]. Four identical fMRI measurements were performed using Simultaneous MultiSlice (SMS) imaging, with six simultaneous slices, a flip angle of 68 degrees, an Echo Time (TE) of 37 ms, and a Repetition Time (RT) of 1000 ms. The matrix obtained was 110x110 voxels with a field of view of 220 mm, resulting in an in-plane resolution of 2mm. The slice thickness was 2mm (66 slices were acquired in each RT). To later correct for susceptibility distortion, field maps spin-echo images were acquired before and between functional runs 2 and 3.

4.3. Stimuli

Nine hundred musical excerpts previously classified into four categories, defined according to Russell's circumplex (labelled as Q1 (positive valence, high arousal), Q2 (negative valence, high arousal), Q3 (negative valence, low arousal), and Q4 (positive valence, low arousal) [10]) were used as the stimuli database in this study [74] .

For each participant, ninety-six musical stimuli of 11.5 seconds were randomly selected: twenty-four stimuli for each quadrant. Within each run, participants listened to 24 stimuli, 6 of each quadrant grouped into two sets of 3 stimuli (each block lasted 36 seconds - 11.5 seconds per stimuli and 0.5 seconds interval between stimuli). Inter-stimuli intervals (ISI) blocks consisted of a 12 seconds component without any sound (excepting the ambient sound inherent to a MRI scanner), followed by 12 seconds with white noise, and ended with a second 12 seconds component with only the ambient sound. The run structure was based on the randomized presentation of Quadrants (two trials) interleaved with ISI blocks. Within each block, all stimuli were presented in pseudo-randomized order so that the second pass of a Quadrant occurred only after the first presentation of all Quadrants. The total duration of each run was 600 seconds.

Participants were asked to perform four music listening runs and were instructed to close their eyes during the runs.

The visual representation of the organization of each trial can be seen in **Figure 13**.

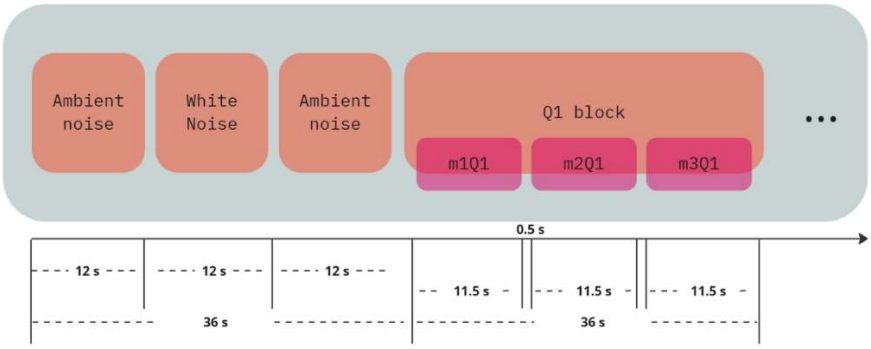

Figure 13. *Visual representation of the organisation of one trial of the experiment.*

After scanning, the participants were asked to classify each song excerpts listened to in terms of arousal and valence, localising them within Russel's circumplex model. Participants were explicitly instructed to assess how they felt. This assessment aimed to obtain a personal perception of the emotional content of each music. The characterisation was defined using a mouse click on a two-dimensional plane –participants were informed that the distance to the centre of the circumplex (neutral valence and arousal) was also important to characterise each music excerpt.

4.4. Data Analysis

4.4.1. Preprocessing

Acquired functional data were preprocessed by members of the research group using fMRIPrep [83]. The fMRIPrep preprocessing pipeline included slice timing correction, motion correction, susceptibility distortion correction using fieldmap images, registration of fMRI data from subject space to template MNI, and estimation of confound signals (CSF, white matter, and grey matter mean time series, framewise displacement, physiological regressors, and six motion parameters).

Data associated to the baseline (ambient sound) blocks was removed. The rationale for removing the baseline was to minimise the impact of auditory primary sensory areas. Since we are exploring the emotional content of music, we define the white noise condition as our within-subject control condition. Then, the data were converted to z-scores.

4.4.2. Feature Selection

MVPA involves the definition of features that characterize and allow the discrimination of the different classes in our classification framework, the classification algorithm, and assessment measures.

Different approaches to feature selection were tested in the initial phase of the work and some of the preliminary results may be consulted in the Supplementary Material (Chapter I – Initial feature selection approaches).

4.4.2.1. Temporal definition of features

Given the duration of our stimuli (each music excerpt lasts 11.5 seconds), we were interested in maximising the emotional response to the stimulus. To isolate the neural instantiation of each music excerpt, the preprocessed voxel activation levels were averaged in a window 6 s after stimulus onset until the end of it (i.e., whole-brain volumes 6 to 12). This choice of the window considered the delay in the hemodynamic response.

4.4.2.2. Spatial definition of features

Finally, regarding the spatial masking, i.e., the spatial definition of the feature set, two methods were chosen to analyse the fMRI images: a data-driven method and a hypothesis-driven method.

Data-driven method: Voxel stability masks

Voxel stability consists of determining the most stable voxels within the training set of each participant and then using those stable voxels as a mask in both the training and the remaining test set (in this work we used the 1000 most stable).

The most stable voxels in the brain are defined as voxels presenting a stable activation profile across the multiple presentations of a set of labels.

Let us consider a 4-class problem defined by the quadrants of the Russell circumplex (i.e., Q1, Q2, Q3 and Q4). The stability of each voxel was computed as the mean pair-wise correlation between its 4 z-scored activation profiles across all pairwise combinations of the multiple presentations in the training data.

Here, the trial-type activation profile of a voxel for a particular presentation refers to the vector of 4 responses of that voxel to each trial type during that presentation. A stable voxel is thus one that responds similarly to the different stimuli set each time the set is

presented. A representative scheme of the calculation of a single voxel's stability is presented in **Figure 14.**

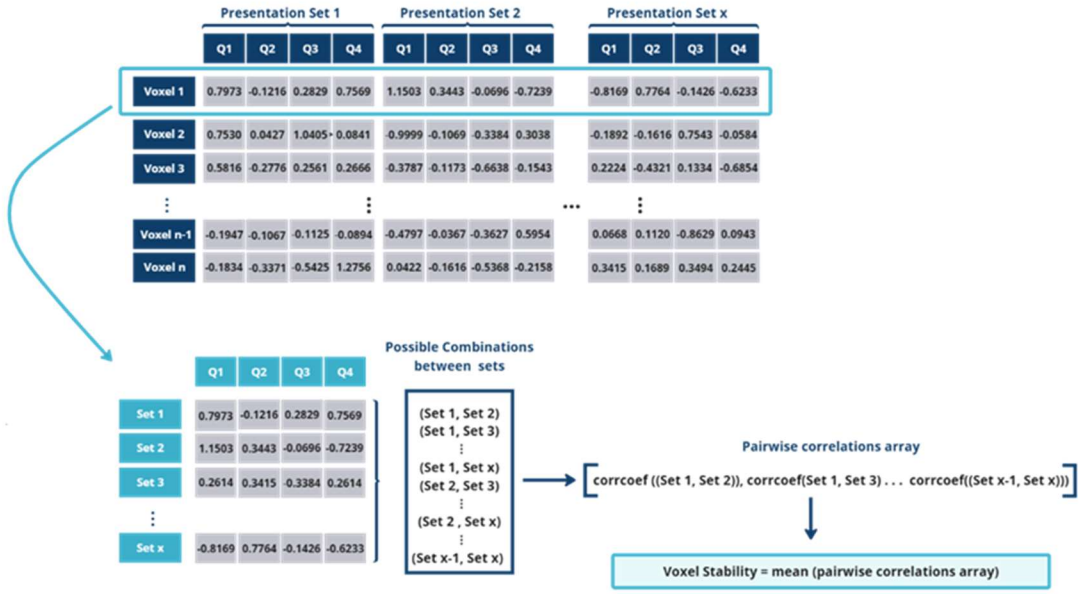

Figure 14. *Schematic representation of a single voxel's stability calculation.*

The pairwise correlation is given by **Equation 10**, where R_{ij} is the correlation coefficient matrix, with values always between -1 and 1, and C is the covariance matrix.

Equation 10. *Correlation Coefficient.*

$$R_{ij} = \frac{C_{ij}}{\sqrt{C_{ii}\, C_{jj}}}$$

Hypothesis driven method: Meta-analysis mask

The second method for spatial constraining of the feature set is a mask defined in a recent meta-analysis by Koelsch et al. (2020) [4] on the neural correlates of music evoked emotions. The meta-analysis examined neuroimaging studies on emotions evoked specifically by music to identify consistent patterns of brain activity across studies. The analysis included 47 studies and found that music-evoked emotions consistently activate a network of brain regions involved in auditory, emotional, and reward processing.

Specifically, the analysis identified consistent activations in the amygdala, hippocampus, anterior cingulate cortex, insula, and ventral striatum/nucleus accumbens.

The statistical image provided by the authors was then converted into a binary mask and applied to all the preprocessed fMRI images. The binary mask obtained is present in **Figure 15**.

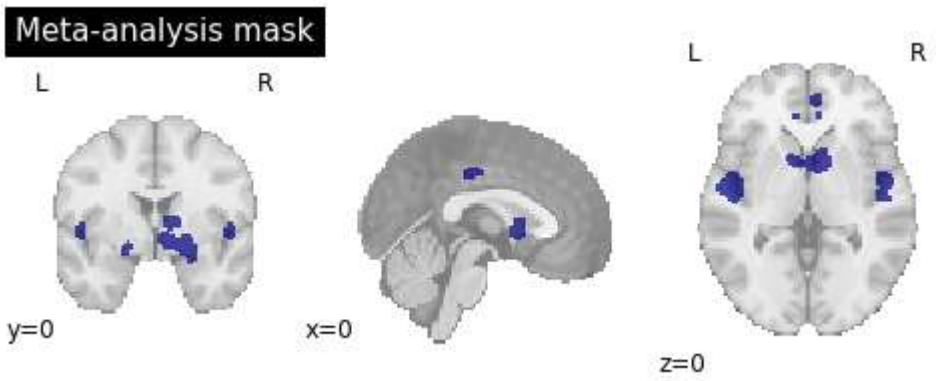

Figure 15. *Mask obtained from the results of the Koelsch's (2020) meta-analysis.*

4.4.3. Class definition

The class definition (i.e., the quadrant to which each music excerpt belonged) was made based on two sets of labels:

i) the original labels associated with each musical clip as defined by Panda et al. (2020) in their MER study, henceforth referred as **PRED** labels (predefined labels).

ii) the labels attributed by the participants to each music clip when they were asked to classify the excerpts after the fMRI experiment, henceforth referred as **PART** labels (classes attributed by the participants).

4.4.4. Training and testing set splits and class imbalance

Data was split into two subsets, training and testing sets. 70% of the data were used in the training set and used to optimise the classification method and the remaining 30% was used to evaluate the model ability to classify the data correctly.

Class definition based on participants labelling often did not match the original labels. This mismatch resulted in a class imbalance where the number of music excerpts in one class could be significantly higher than others. To address this class imbalance, we tested two solutions: to maintain the class balance, limiting the training set to the number of samples with the class with the least data points and to oversample the class with the least data points.

Limiting the training set and restricting the number of data points resulted in poor training performance for the classifiers due to the limited number of samples.

The oversample of the classes with a lower number of data points was used in the training set to ensure that all participants had at least 12 data samples of each trial type.

Oversampling is a method used to increase the number of samples in the minority class to match a desired number of samples. In this case, the Synthetic Minority Over-sampling Technique (SMOTE) [84] was utilized. SMOTE generates synthetic samples for the minority class by linearly interpolating between existing samples and their k-nearest neighbours.

Using the SMOTE algorithm, a new synthetic sample, x_{new}, is generated by interpolating between an existing sample, x_i, and one of its k nearest-neighbours, x_{zi}), using a random number λ within the range [0, 1]:

***Equation 11**. Generation of a new sample with SMOTE.*

$$x_{new} = x_i + \lambda \times (x_{zi} - x_i)$$

4.4.5. Decoding analyses

Linear-kernel support vector machine classifiers were trained to perform four different decoding analyses:

- music vs. noise
- positive vs negative valence level
- positive vs negative arousal level
- Both the valence and arousal levels of a specific music excerpt simultaneously, i.e. Q1, Q2, Q3, Q4.

The first 3 classifiers represent binary classification problems, while the last is a multi-class problem. In the multi-class problem, the solution is obtained in a 'one-vs-all' approach (for each class, a classifier if fitted against all the other classes). All classifiers were trained in a training set to find the optimal number of voxels to use in the stability mask and to optimize the margin-parameter, C, in a grid-search 5-fold cross validation. After training, the optimal parameters were used to do predictions on the last 30% of the data (test set).

4.4.5.1. Kernel selection

The choice of the classifier considered SVM's with both RBF and linear kernels, according to examples from the literature ([74] and [65], [85], respectively). Since in the preliminary tests there was no significant difference between the performance of the classifiers when using both types of kernels (for the preliminary results, please refer to the Chapter 2 of the Supplementary Material - Comparison of SVM classifier with linear vs RBF kernel), the linear kernel was chosen. This also allowed us to obtain the features contribution to the classification problem and explore the main brain areas relevant for the classification of each type of stimuli.

4.4.5.2. Feature set and class definition

Additionally, the classifiers were trained using both the PRED and the PART labels, resulting in eight classifiers for each of the two feature selection methods. In total, sixteen classification models were created and the combinations between feature selection method, labelling strategy and decoding analysis are presented in **Figure 16**.

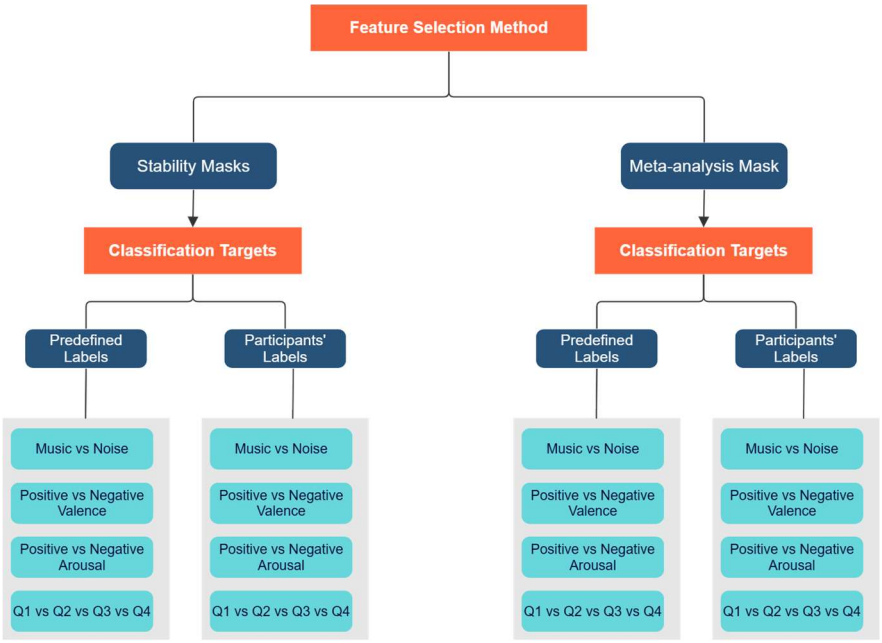

***Figure 16**. Characteristics of each of the sixteen classification models created.*

All analysis scripts were developed in *Python 3.9.12*, utilising the Anaconda distribution, and executed and documented using *Jupyter Notebooks*.

4.4.6. Statistical Significance

Permutation Tests

To assess whether the results provided by each of the classifiers were statistically significant, permutation tests (i.e., randomly permuting the labels of the dataset while keeping the features unchanged and then calculating a score metric) with 1000

permutations were employed in the test sets to compare the obtained results to a null distribution and assess their significance. The permutation tests were performed considering the accuracy as the performance metric.

An empirical value, p, is then calculated as the proportion of permutations for which the score obtained is equal or greater than the score obtained using the original data [86], [87].

In this specific case, a threshold of 0.05 was for p was used to determine statistical significance, which means that if the $p<0.05$, it suggests that the observed accuracy is unlikely to occur by chance alone, and the null hypothesis of no dependency between features and labels can be rejected, i.e.: there is evidence of a relationship between features and labels and the performance of the classifiers can be considered statistically significant.

Exploring differences between approaches

We also explored the best configuration of features selection methods and classification pipeline, comparing the results between different approaches and determining the statistical significance.

To assess if the data followed a normal distribution, we used the Shapiro-Wilk test due to its widespread usage and higher statistical power, particularly for small sample sizes ($N<50$) [88]. For details about the implementation of this method please refer to the original paper, [89].

To compare the classification results, when these followed a normal distribution, a paired student t-test was used. This is the most common parametric method and compares the means of two different sets of data to determine if they are equal; if they are, then no difference exists between the sets. When the data did not follow a normal distribution, the non-parametric equivalent to the student t-test, the Wilcoxon signed rank test that analysis the median of the two populations [90].

4.4.7. Identifying brain areas contributing for the decoding

After classification, the brain areas that were the most relevant for the classification process were determined.

To do this, we first performed an inversion of the SVC coefficients obtained in the classification step, recreating a statistical image for each classifier and each of the participants. The final statistical images were the result of the combination of the 100 most significant voxels of each participant for each of the four decoding models (in the multiclass problem, a statistical image for each of the four quadrants was obtained and the contributions of the voxels to each of the classes were combined in a single image).

The coordinates of the statistical clusters were then obtained in the MNI 152 space.

Finally, the identification of the brain areas that corresponded to each of the found clusters was done utilising an adapted python script by Astrid Olave of the original MATLAB code from Xu Cui (https://www.alivelearn.net/?p=1456).

This script utilises the Automatic Anatomical Labelling (AAL) atlas to identify the anatomical structures that corresponded to the provided coordinates.

5

Results

This chapter presents the results obtained in the present work. Section 5.1 shows the results of the behavioural task and differences between labels according to the two strategies considered (PRED *labels and* PART *labels*), Section 5.2 evaluates the influence of labelling strategy on feature selection and 5.3 details the results on the four decoding analyses.

5.1. Behavioural categorisation task

To illustrate the differences between the two labelling strategies, we present the confusion matrix considering the two types of labels in **Figure 17**.

We found a high discrepancy between the two labelling approaches.

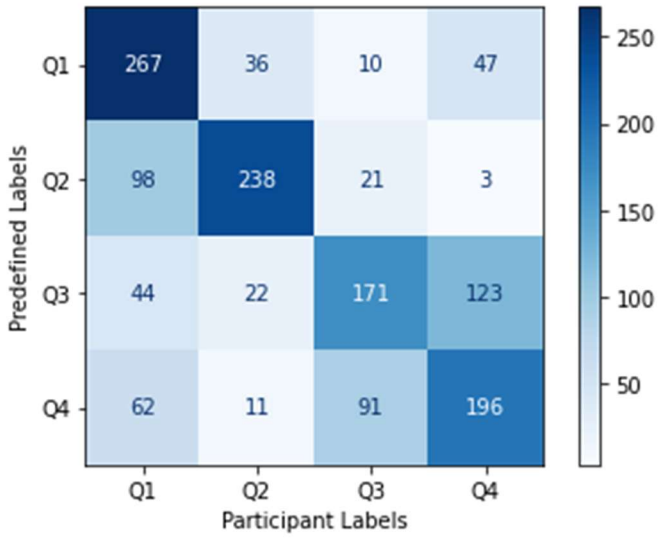

Figure 17. *Confusion Matrix comparing the PRED labels with the PART labels.*

There were 360 music clips from each quadrant in total. The confusion matrix shows that participants preferentially labelled musical excerpts as having both positive valence and

arousal, i.e., as belonging to the first quadrant, Q1, over the other categories (the participants attributed the label 'Q1' to a total of 471 music clips, contrasting with a mean of 323 attributions to each of the other labels).

Additionally, the overall mismatch between the tags attributed considering the two types of labelling is particularly evident in the quadrants Q3 and Q4, which are often re-labelled.

5.2. Preprocessing and feature selection

We defined two approaches to select voxels subsets, to address the potential issues associated with the curse of dimensionality. We next present the results of the stability measure.

5.2.1. Stability Masks

The stability masks were determined according to the method presented in 4.4.2.2.

The following results show an example of the obtained masks for a single participant in all four possible situations (in the two labelling strategies considering both five and four classes in the creation of the stability mask).

Figure 18.A presents an example of the voxel stability map for the 5 class model (Q1, Q2, Q3, Q4 and noise) and PRED labels of one single participant, and **Figure 18.B** presents the voxel stability map for the 5 class model with the PART selection of labels of the same participant.

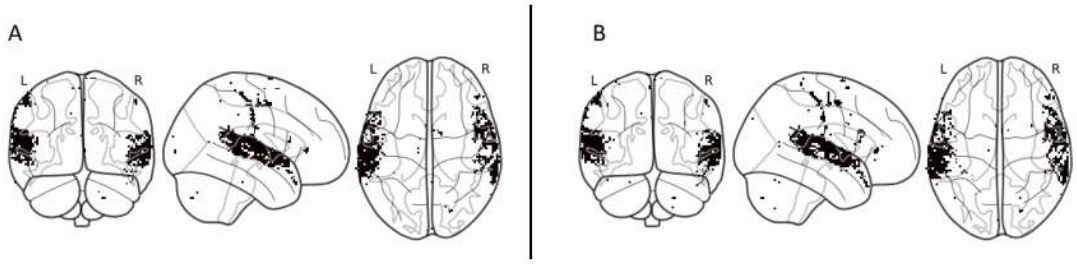

***Figure 18**. Stability masks with 5 classes of participant SUB-07. **(A)** Mask created considering PRED labels. **(B)** Mask created considering the PART labels.*

The masks obtained using the participants' and predefined labels exhibited consistent results for the 5 classes situation, indicating that the most stable voxels were located in comparable brain regions, particularly around the auditory cortex, regardless of how the stimuli were subjectively perceived.

Figure 19.A presents the voxel stability map for the 4-class model (Q1, Q2, Q3, Q4) and PRED labels, and **Figure 19.B** presents the voxel stability map for the 4-class model with individual selection of labels (PART).

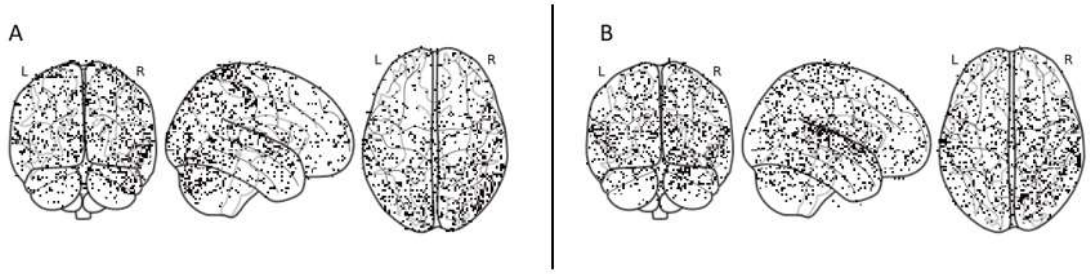

Figure 19. *Stability masks with 4 classes of participant SUB-07.* ***(A)*** *Mask created considering PRED labels.* ***(B)*** *Mask created considering PART defined labels.*

When considering the voxels' activation profiles for the four quadrants only, without noise, the consistent pattern around auditory areas verified before was replaced by a noisy random distribution of voxels.

5.3. Decoding Analysis

In this section, we present the results for the four different classification pipelines (two labelling strategies (predefined and tagged by the participants) and two feature selection masking strategies (stability masks and the mask retrieved from Koelsch et al. (2020) meta-analysis).

The performance was assessed considering the training score of the best classifier (i.e., the mean accuracy across all cross-validation folds with the best model parameters), its

overall accuracy in the testing set, and the value of p that resulted from the permutation tests performed for each model to determine the statistical significance of the results. We define as lower threshold for the significance of a classification model, as one that present statistically significant results ($p<0.05$) for at least 50% of the participants.

The detailed classification results for each participant can be consulted in the Supplementary Material (3. Discriminated results for each individual label).

The plot bars present the results for the mean performance across all participants.

5.3.1. Predicting music and noise

The first objective was to assess the ability of decoding periods of listening to musical excerpts and periods of listening to white noise. The results of this task are presented in **Figure 20**.

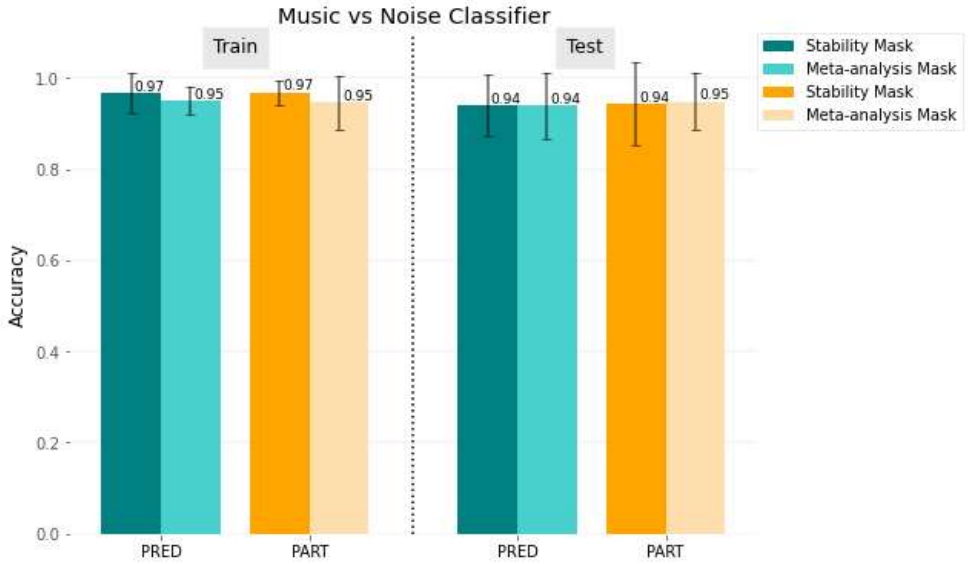

***Figure 20**. Performance of the model classifying music and noise*

The results show that when considering either the PRED labels or the PART labels, the classifier successfully distinguished between the two types of stimuli despite the feature selection method used. The permutation tests conducted on the data showed that the classification performance was statistically significant ($p<0.05$) for all participants in the four

situations. This indicates that the classifier could differentiate between the auditory experience of listening to song excerpts and white noise.

No significant difference was found when comparing the two sets of labels (original vs. participant-provided) using the Wilcoxon Rank test ($p>0.05$).

Moreover, when using the meta-analysis mask, there is a small decrease in the performances for both types of labels.

In the forthcoming results, since the classification model did not include the decoding of Noise volumes, the considered stability mask is always the one created with 4-classes only.

5.3.2. Predicting positive and negative valence

Our second objective was to decode the valence of musical stimuli: classify music clips with positive valence (Q1 and Q4) and clips with negative valence (Q2 and Q3). The performance results when using the two feature selection methods and both types of labels are presented in **Figure 21**.

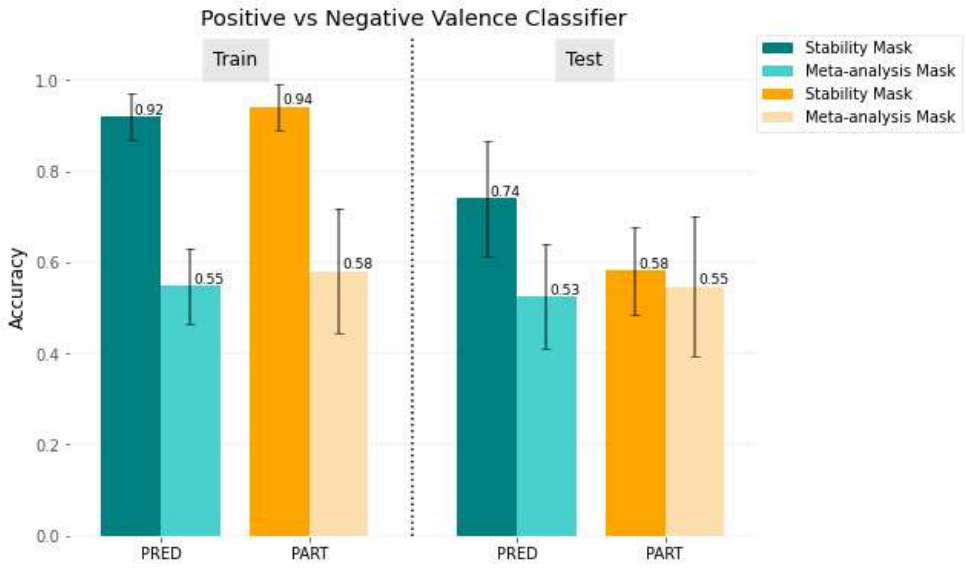

***Figure 21**. Performance of the valence decoding model.*

When considering the PRED labels in the data masked with the stability mask, the results (74%±13% accuracy in the test set) are substantially superior to the chance level (50% for a balanced dataset). Similarly, to what was verified by the music vs. noise classifier, the performance is always superior using stability masks over the meta-analysis mask. Even within the training sets, there is an even more evident discrepancy between the mean accuracies. Moreover, the results suggest some overfitting, particularly in the stability masking strategy (for both sets of labels), as the results using the training set are far superior to those obtained in the testing set. In this sense, the data demonstrate that the classifier can correctly identify patterns in the features with these masks but has poor generalization ability.

Training results using the meta-analysis mask are close to the chance level, suggesting that the model cannot learn a decision function able to separate the two classes.

Additionally, when comparing the utilization of the two types of labels as targets of the classification, there is a significant difference ($p<0.05$) between the PRED and the PART (58%±10% accuracy in the test set) labels with the stability masks.

However, this difference between the test results for the two types of labels is not verified when the meta-analysis mask is used (overall accuracies in the test sets of 53%±11% and 55%±16% considering the PRED and the PART labels, respectively, and $p>0.05$). This result is not interpretable due to the poor performance of the classification model.

Finally, it is important to consider the permutation test results. Notably, the classification results for the combination of the PRED labels with the stability masks were significant, contrasting with a lack of significance in all the other classification models, not allowing for an interpretation of those results.

5.3.3. Predicting positive and negative arousal

The third model's aim was to decode arousal. In this sense, the classification problem was separating between data points referring to music with high (Q1 and Q2) and low arousal (Q3 and Q4). Similarly, to the previous models, the performance results in terms of mean

accuracies for all possible combinations of feature selection and targets in training and testing are presented in **Figure 22**.

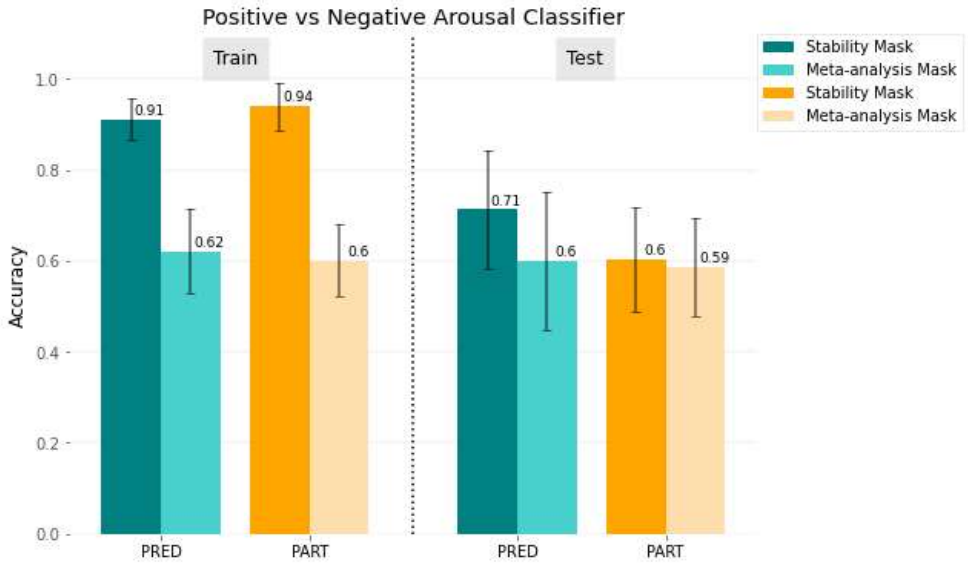

Figure 22. *Performance of the arousal decoding model.*

The best results were obtained with the stability masking strategy and PRED labels (71%±18% accuracy in the testing set). When considering the PART labels (60%±10%), there is a significant decrease ($p<0.05$ for the paired t-test between the two sets of labels) in the performance of the model.

As in the previous analysis, when the meta-analysis mask is used, the results in the training and test (60%±16% with the PRED and 59%±11% using the PART labels) sets are very similar for both types of labels. This result is not interpretable due to the poor performance of the classification model.

5.3.4. Predicting each individual quadrant

The final model objective was to identify each of the four quadrants, a multi-class problem. Chance level in this multiclass problem is 25% in the predefined label selection (the individual selection of labels introduces some variability). The results for both types of labels and feature selection approaches are presented in **Figure 23**.

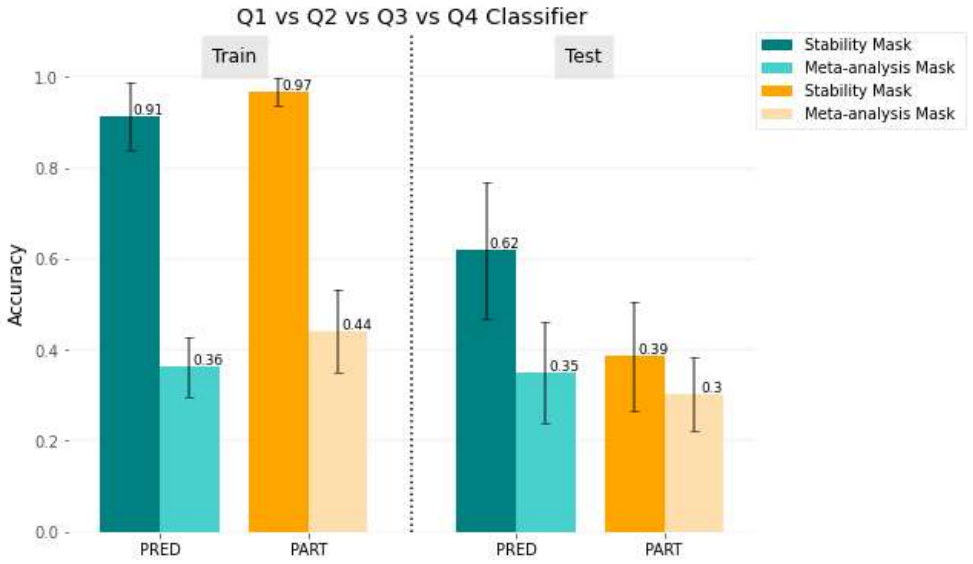

Figure 23. *Performance of the individual quadrants decoding model*

The best results were obtained using the stability mask as feature selection strategy and PRED labels (testing set accuracy of 62%±15%). When comparing the two feature selection approaches, the use of the stability masks revealed, once again, a significantly higher performance ($p<0.05$) than the meta-analysis mask (accuracy in the test sets of 35%±11% and 30%±8% considering the PRED and the PART labels, respectively)

Confusion Matrices

To further characterise the prediction pattern of this model, we present the normalised confusion matrices (the values represent the percentage of correct classifications for a specific class), associated with each of the four possible combinations between masks and targets used.

First, we show the matrix for the stability mask feature set (**Figure 24**).

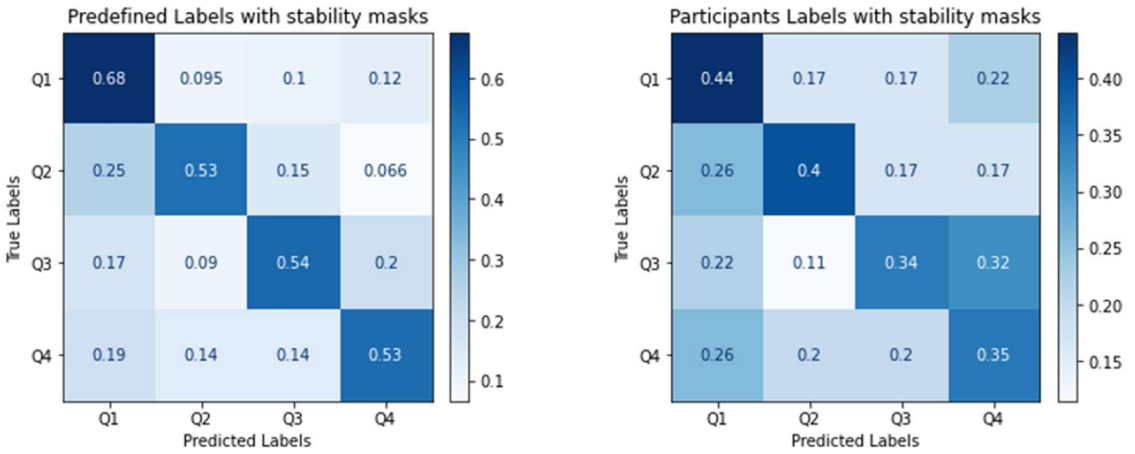

Figure 24. *Confusion Matrices when stability masks were used as feature selection method for both types of classification labels.*

The results show that the model with the predefined labels tend to correctly identify the target label and no pattern emerge regarding the false negative. Considering the participant's labelling confusion matrix, the results show a decrease across all classes. Again, no particular pattern arises from the data regarding missed predictions.

Figure 25 shows the meta-analysis masking strategy results.

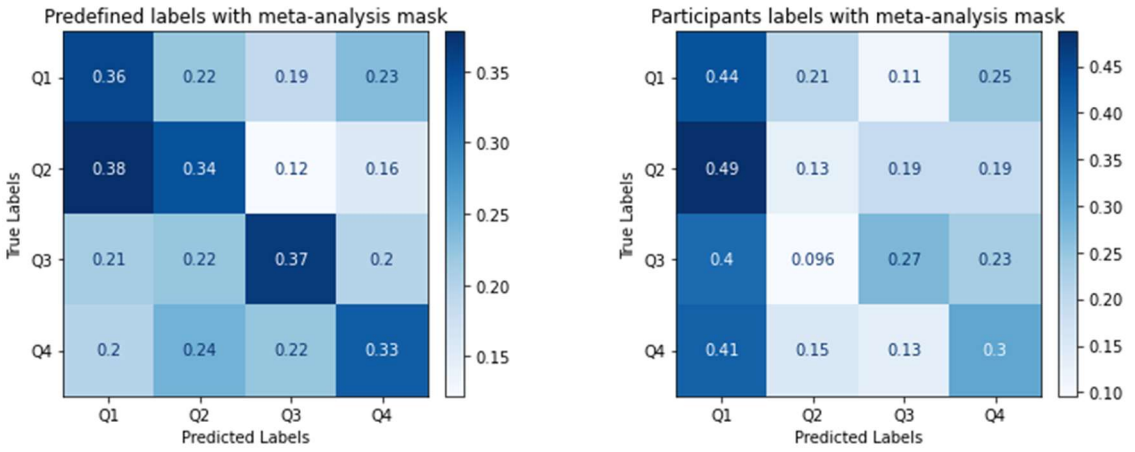

Figure 25. *Confusion matrices when the meta-analysis mask was used as a feature selection method for both types of classification labels.*

The results show that the model with the predefined labels tend to correctly identify the target label (inferior results compared to the previous feature selection strategy) and no pattern emerge regarding the false negative distribution. The classifier using PART labels as targets together with a meta-analysis mask shows a clear bias to predict Q1.

5.3.5. Significant brain regions for decoding

To further characterize the successful models, we detail the results for the combination of PRED labels as targets of the classification and the stability masks as the feature selection method (as the other models presented poor training results, and therefore their interpretability is limited). To this end, we defined masks of interest based on combining each participant's 100 most significant voxels for each decoding model. We then present the histogram of the number of clusters found in the ten cortical and subcortical regions with the highest number of clusters.

5.3.5.1. Music vs Noise

The location of the most significant voxels used by the model to distinguish between music and noise are presented in **Figure 26** and the number of clusters found in each area is shown in **Figure 27**.

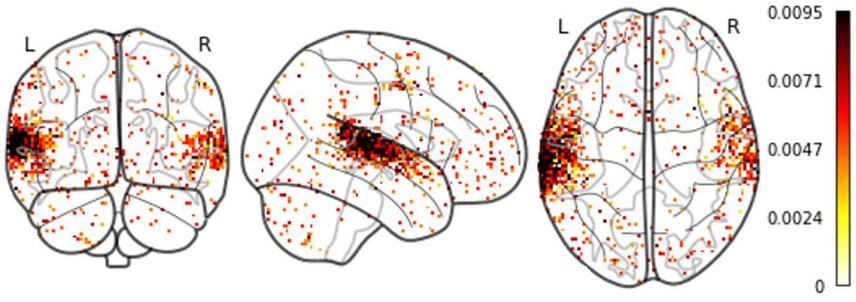

Figure 26. *Location of the most significant voxels in the music vs noise classifier. Voxel values correspond to the combination of the SVC coefficient values of that voxel.*

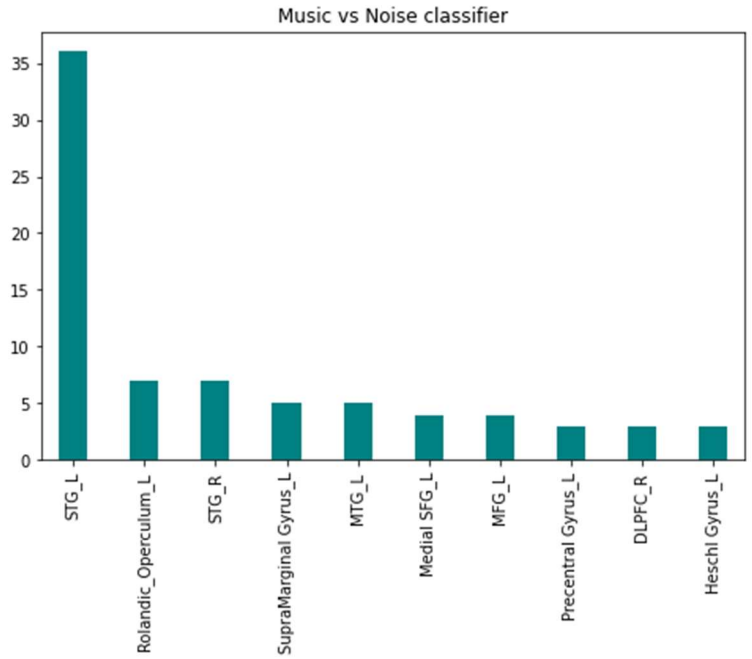

Figure 27. *Distribution of the significant clusters for the decoding of music and noise across the top 10 regions with the highest number of clusters. STG: superior temporal gyrus; MTG: middle temporal gyrus; SFG: superior frontal gyrus; DLPFC: dorsolateral prefrontal cortex; R: right lobe; L: left lobe*

The significant voxels for decoding music and noise were distributed across the entire region outlined by the stability mask created using the 5 classes. The majority of these voxels were situated within the auditory cortex, particularly across the STG and transverse gyrus (the cluster coordinates in the Rolandic operculum correspond to the auditory cortex according to [91]. Additionally, a few smaller clusters were identified in the frontal (DLPFC and precentral gyrus) and in the parietal (supramarginal gyrus) lobes.

5.3.5.2. Positive vs negative valence

In **Figure 28** and in **Figure 29** the distribution of significant clusters across the brain for the decoding of positive and negative valence is presented.

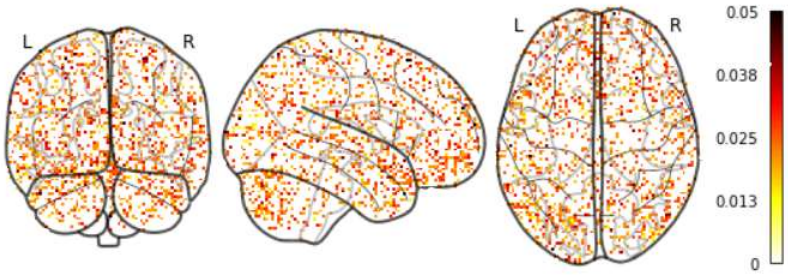

Figure 28. *Location of the most significant voxels in the positive vs negative valence classifier. Voxel values correspond to the combination of the SVC coefficient values of that voxel.*

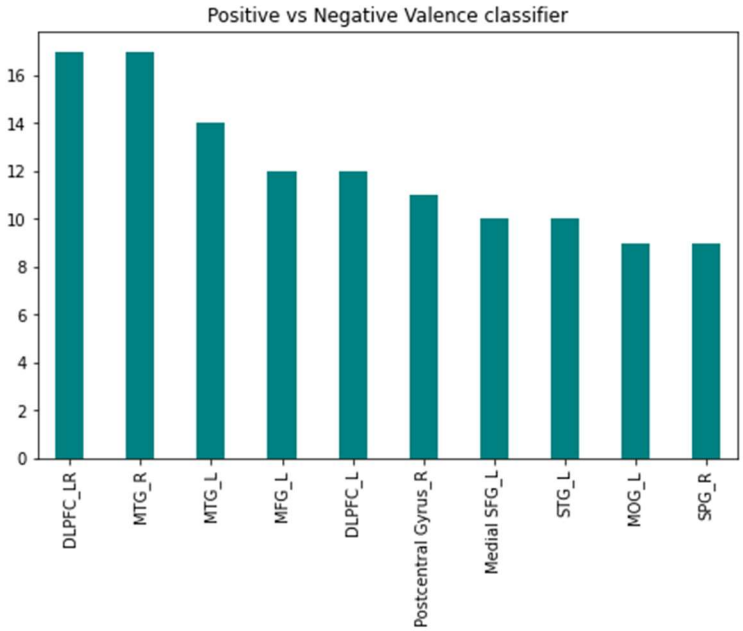

Figure 29. *Distribution of the significant clusters for the decoding of positive and negative valence across the top 10 regions with the highest number of clusters. DLPFC: dorsolateral prefrontal cortex; MTG: middle temporal gyrus; MFG: middle frontal gyrus; SFG: superior frontal gyrus; STG: superior temporal gyrus; MOG: middle occipital gyrus; SPG: superior parietal gyrus; R: right lobe; L: left lobe*

5.3.5.3. Positive vs negative arousal

The significant clusters for the decoding of positive and negative arousal are shown in **Figure 30** with the correspondent distribution of the number of clusters found in each area presented in **Figure 31**.

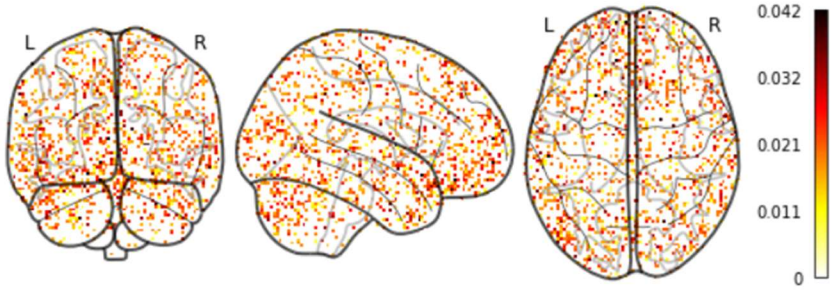

Figure 30. *Location of the most significant voxels in the positive vs negative arousal classifier. Voxel values correspond to the combination of the SVC coefficient values of that voxel.*

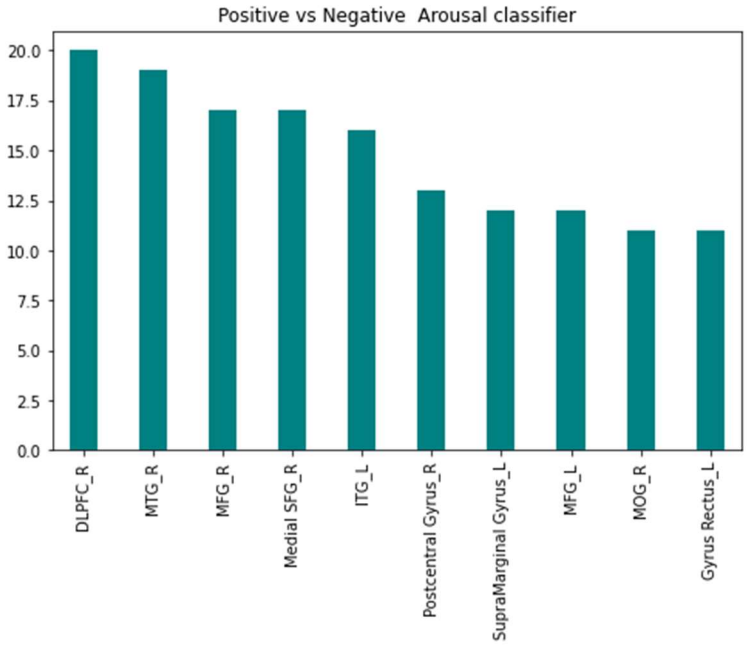

Figure 31. *Distribution of the significant clusters for the decoding of positive and negative arousal across the top 10 regions with the highest number of clusters. DLPFC: dorsolateral prefrontal cortex; MTG: middle temporal gyrus; MFG: middle frontal gyrus; SFG: superior frontal gyrus; ITG: inferior temporal gyrus; MOG: middle occipital gyrus; R: right lobe; L: left lobe*

5.3.5.4. Q1 vs Q2 vs Q3 vs Q4

Finally, we present the distribution of voxels with significant information for classification of each individual quadrant in figures **Figure 32** and **Figure 33**.

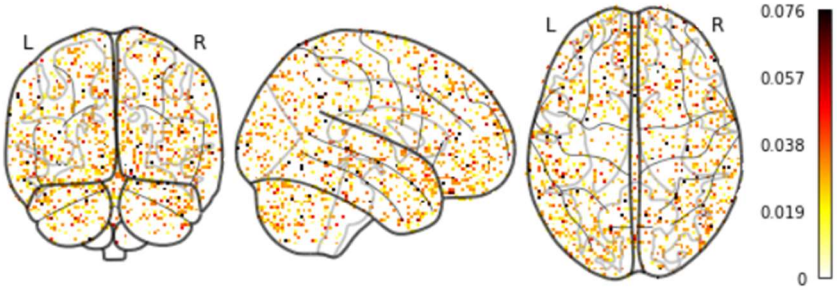

Figure 32. *Location of the most significant voxels in the classifier decoding each quadrant individually. Voxel values correspond to the combination of the SVC coefficient values of that voxel.*

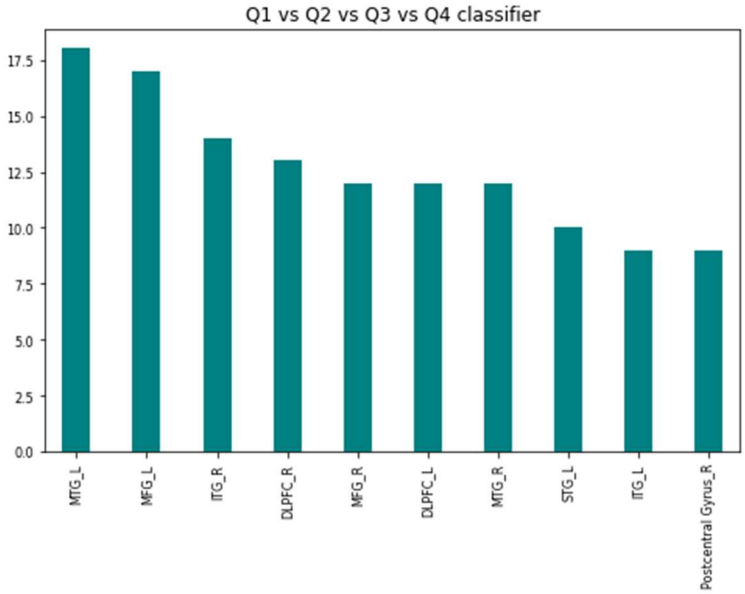

Figure 33. *Distribution of the significant clusters for the decoding of each individual quadrant across the top 10 regions with the highest number of clusters.. MTG: middle temporal gyrus; MFG: middle frontal gyrus; ITG: inferior temporal gyrus; DLPFC: dorsolateral prefrontal cortex; STG: superior temporal gyrus; R: right lobe; L: left lobe*

The distribution of the voxels containing significant decoding information when using stability masks created only from the voxel activation patterns in response to 4 classes replicates the noisy patterns observed in these masks, with small clusters distributed across the whole brain. The results are very similar for all the three classification models.

Regarding the information in the histograms, the majority of the decoding information was located in neocortical regions. The results meet previous research using MVPA to decode feeling states, where neocortical regions were identified as encoding emotional information in the brain [67], [68], [70]–[72].

Most of the clusters were found bilaterally in the auditory cortex (STG and Heschel's gyrus), in visual areas in the temporal (MTG, inferior temporal gyrus (ITG)) and occipital lobe, as well as is in prefrontal cortex (DLPFC, ventromedial prefrontal cortex (VMPFC), middle frontal gyrus (MFG)) and across the occipital, parietal (precuneus) and temporal lobes (fusiform gyrus, MTG), as well as in the primary somatosensory cortex (postcentral gyrus).

Additionally, some clusters were found in a smaller number in premotor areas (precentral gyrus), parietal regions (precuneus, superior parietal gyrus, SPG) and in the cingulate cortex.

Also, although the voxels with significant information for the decoding were mainly identified in cortical areas, some subcortical regions were also found in less number. In particular, the hippocampus, the amygdala and the insula revealed to contain important in decoding information. These regions are known to be involved in emotional processing and have been identified in multiple studies as relevant for the processing of specifically music-evoked emotions [4], [67], [68].

However, contrarily to other decoding studies, no significant voxels were identified in the parietal operculum, more specifically, on the secondary somatosensory cortex. This area has been shown to be engaged during multiple experiences of perceived emotions and, in particular, to take an important role in music-evoked emotions [67], [68].

6

Discussion

6.1. Defining the decoding target

The behavioural categorisation task results evidenced a discrepancy between the different attribution of labels. This mismatch was particularly evident in the labels from Q3 and Q4 which were often relabelled for each other, revealing increased inter-individual variations in the definition of positive and negative valence for low arousal stimuli compared to the PRED labels.

This relabelling pattern concurs with what was verified by Panda et al. (2020) [74]. In their study, where the predefined labels were set: their participants also manifested more difficulties in classifying low arousal music, which can explain the differences in the labelling present in our confusion matrix.

6.2. Exploring the optimal feature set and model parameters

The present work aimed to create models that identified brain patterns in fMRI images, allowing them to distinguish between: music and noise, positive and negative valence, positive and negative arousal and between each of the individual Russell's quadrants simultaneously.

All models trained to decode music vs. noise achieved high accuracy and statistically significant results. As this is a binary classification problem (all music excerpts belong to the same class, i.e., music), the labelling strategy does not influence the model training. Considering the feature selection method, both methods achieved high accuracy. In this

sense, for the music vs. noise problem, the specific constraining of voxels within the two different masks does not influence the performance of the classifiers. These results may be related to the fact that both the stability masks and the meta-analysis mask included voxels in temporal regions across the STG. Discriminative patterns for music were found to be widely distributed bilaterally in these areas in a study decoding multiple sound categories [72].

On average, the decoding of positive and negative valence was above chance level (considering 50% for a balanced dataset) for all the pipeline combinations. However, only the combination of stability masks with the PRED labels presented results above the significance threshold. Therefore, we can consider that it was possible to decode the valence level of a music excerpt by finding patterns in the fMRI images using that model. The results for all the other models are not interpretable due to the lack of significance.

The results of the arousal classification, even though they were also above chance level for all the models, were not statistically significant for any of them (did not surpass the considered significance threshold), suggesting that their performances may have been a result of random attribution of labels instead of the correct generalisation of the patterns learnt in the training phase.

Notably, in the decoding of both valence and arousal, there was an overfitting of the data solely when the stability masks were used, evidencing that the models found problems only in the generalisation of the learnt patterns to unknown data. The masking of the data with the meta-analysis mask supressed the learning of patterns even in the training phase.

Regarding the multi-class models, the best significant results were again obtained when using the PRED labels as targets of the classification with the stability masks as the feature selection method (accuracy of 62%±15% in the testing set).

The false predictions were distributed across all classes, not following a specific pattern, contrarily to what was verified in other studies. In particular, Panda et al (2020) [74], (that used musical features of the same *stimuli* dataset to decode each quadrant), verified music from certain quadrants was more often misclassified as belonging to other quadrant in specific. This suggests that the link between the musical features associated with certain

emotions may not correlate with the neural activation patterns associated with those same emotions, these being highly dependent on the context and on the individual perception of the stimuli.

6.3. Feature selection methods

The stability masks provided better classification results in all the decoding tasks. Even in the models with lower accuracy in the testing set, the classifiers presented high accuracies in the training set, indicating that they were able to learn patterns in the data despite the lower generalisation ability (this is particularly interesting since the performance in the training set was assessed using cross-validation). Conversely, when the meta-analysis mask was applied, the models could only discern between music and noise. For all other decoding tasks, they exhibited an inability to acquire meaningful patterns from the training dataset or to extrapolate knowledge to unfamiliar data points. The fact that the meta-analysis mask allowed to classify music and noise, suggests that this selection of voxels is enough to distinguish a musical auditory stimulus from a non-musical, but it does not identify the specificity of its emotional content.

This indicates that, while the regions encompassed by the meta-analysis may be indicative of broad emotional processes, they might not capture the individual and context-dependent neural signatures that distinguish music evoked emotions as effectively as the stability masks, which are tailored to each individual participants' brain activation patterns.

Furthermore, the stability masks were generated using training sets comprising volumes from various music clips associated with the same emotional quadrant, without repetition across functional runs. This approach allowed the classifiers to learn patterns related to general valence and arousal levels, rather than specific songs. As a result, the models exhibited enhanced generalisation capabilities, as they did not focus on the idiosyncrasies of individual songs.

On the contrary, many of the studies used to create the meta-analysis mask concentrated on a limited range of emotional labels, (e.g., joy and fear), and employed a

relatively small stimuli dataset. This narrower scope could potentially limit the ability of the meta-analysis to capture the full diversity of music-induced emotional experiences.

In fact, the poor performance of the models when considering the meta-analysis mask correlates with the current decoding studies that either did not find voxels with significant information for the decoding of emotions within these regions [67], [68], [71], or obtained poor classification performances across them [92].

This might highlight the inherent variability in how different individuals' brains encode emotional states and the limitations of generalising from meta-analytic findings to individual-level emotion classification tasks.

6.4. Class definition methods

The PRED labels originated better results compared to the PART labels when the stability masks were used as feature selection method. However, when the meta-analysis mask was used, no significant difference was found between the use of each labelling strategy. Nonetheless, the absence of differences between the class definition methods results is not interpretable for this second type of masking due to the poor performance and significance of these models.

It is possible that, when the stability masks were used, the differences between the types of labels used as targets of the classification were a consequence of the imbalance introduced by the relabelling of the dataset. This imbalance was addressed by increasing the number of points of the classes with least data points in the training set so all the classes would have the same number of points. However, in order to preserve the optimal representation of a real-life problem, this oversampling step was performed only in the training set, resulting in uneven testing sets. In particular for the multi-class problem, when the labels attributed by the participants were used, there were situations where a participant had only one or two points of a certain quadrant and more than ten of the others. As consequence, in these situations, if the classifier failed to classify a single point, that would cause a large decrease in the performance of the model when classifying that class.

The fact that the differences in the results between both types of labels are only present in the testing sets (where the number of points of each label is very discrepant) corroborates the idea that this might be a consequence of the imbalanced and limited test sets and not a reflection the subjective perception and interpretation of the stimuli influence in the individual brain-patterns that arise from music-evoked emotions.

6.5. Brain regions with significant voxels for the decoding

The results observed when stability masks were used are comparable to the results of previous decoding studies regarding music-evoked emotions, where neocortical regions were able to predict the emotional content of a song but the limbic regions were not able to reveal emotion-specific activity patterns.

Regarding the information found in the temporal lobe of the brain, the auditory cortex (i.e.: STG and Heschel's gyrus) is known to have anatomical connections with multiple limbic and paralimbic structures involved in the generation of affective activity, such as the amygdala, the orbitofrontal and the cingulate cortices [93]–[95]. Therefore, it has been previously associated with functions beyond the mere perception of auditory *stimuli:* it has been identified in both decoding ([68], [70]–[72], [96]–[98]) and contrast studies ([54], [66], [99]) as a crucial region for the processing of the emotional content of auditory *stimuli* and of non-musical emotions [92]. Similarly, the temporal poles are highly interconnected with both the amygdala and orbital frontal cortex and have been linked to various functions related with emotion and social cognition, in particular with the emotional processing of auditory *stimuli* [100]. In fact, an hyperactivation of the TPs when a subject is listening to unpleasant music, contrasting with a deactivation when the individual is listening to music with a positive valence has been shown [54].

Nevertheless, we cannot exclude the possibility that the contributions of voxels in the auditory cortex were driven, at least in part, by acoustical differences between stimuli.

Furthermore, somatosensory and motor regions such as the premotor cortex, the SMA and the pre- and postcentral gyri (primary motor and somatosensory cortices, respectively) have been proposed to constitute an interconnected neural circuit with auditory areas, crucial for the perception of music [101] and the presence of significant voxels in these areas correlates with previous studies where these regions also provided above-chance classification accuracies of music-evoked emotions [68], [96].

The involvement of these motor areas in the perception of music might be explained by music's rhythmic entrainment, which refers to the synchronisation of neural activity with the music's tempo, activating motor areas. That is, even in the absence of overt movement, it seems to exist a perception-movement coupling. This also relates with the concept of motor imagery, which refers to the mental simulation of a movement without any actual execution of the movement [102]. Moreover, the pre-SMA area, in particular, is known to be involved in the inhibition of movement in certain contexts, which may happen when the participants go against the urge to move while in the fMRI scanning process. Considering that the urge to move may vary according to the type of musical *stimuli* being present, this brain region may provide significant information about the emotional content of a song.

Additionally, clusters were also found in both brain hemispheres in the DLPFC, both in the *pars triangularis* and in the *pars orbitalis* of the IFG, and in the VMPFC, (in the medial, posterior and anterior orbital *gyri* of the orbitofrontal cortex).

These results are consistent with the critical role of the PFC in the reward network of the brain and in the generation and regulation of emotions [103] as evidenced in previous studies, including music-evoked emotions studies [67], [68], [104], [105].

In particular, the VMPFC is often recruited during the processing of self-related information and autobiographical memories [106], [107], as well as introspection [108] and mind wandering [109], which are processes often related to the arising of emotions. In regard to music, specifically, it has been associated with the subjective perception of the valence of songs[107] and, in particular with high valence and low arousal music) [110] as well as with the emotive processing of unexpected musical chords [111].

Furthermore, the VLPFC and the DLPFC are also involved in the regulation of emotional experiences in terms of valence and arousal [112]. The dorsal medial prefrontal cortex (DMPFC), which is adjacent to the DLPFC, is involved in associating music with memories, particularly emotionally salient episodic memories, suggesting a role of the DLPFC in the encoding and retrieval of music-evoked autobiographical memories [107]. Additionally, activations in the VLPFC were associated with aversive and dissonant music [105] and therefore this area may be used to distinguish between some of the quadrants with negative valence (usually associated with dissonant music).

Furthermore, the presence of significant voxels in visual regions spanning the entire occipital lobe (including the middle, superior, and inferior occipital gyrus, as well as the lingual gyrus and the calcarine fissure with its surrounding visual cortex), in conjunction with temporal regions such as the fusiform gyrus and the MTG, as well as in the precuneus, corroborates earlier findings ([96], [110], [114], [115]). These observations can be explained by visual imagery serving as a mechanism through which music can stir emotions.

In particular, sad music (usually associated with both low arousal and low valence, quadrant Q3 of the Russell's Circumplex) has been demonstrated to be associated with stronger mind wandering through visual imagery in comparison with happy music (high arousal and valence, quadrat Q1 of the Russell's Circumplex), which suggests a role of these visual areas in distinguishing between these quadrants during classification tasks [115].
Additionally, the presence of significant voxels in the precuneus might not only signify its involvement in visual imagery but also underscore its role in episodic memory. This region is intricately linked not only to the creation of mental visual scenarios but also to the retrieval of personal experiences. [116] Furthermore, the parietal lobe unveiled additional clusters within memory-related regions, particularly in the inferior parietal gyrus encompassing both the angular and supramarginal gyrus.

Notably, the left angular gyrus has a key role in episodic memory [117], in particular, in the processing of semantic musical memory (the ability to recognise familiar songs or melodies by their tune, without necessarily remembering the specific context in which they were heard) [118].

Considering the subcortical regions, the amygdala and the hippocampus are central hubs of emotion processing and key elements of the reward network of the brain, interconnected with multiple other regions involved in the processing of feeling states, including the NAcc, the PFC and the auditory cortex. [105] This centrality has been reflected in many studies showing activations in these area during the listening of music with emotional content.

Moreover, whilst the relevance of specific brain regions in the processing of emotions, particularly in the context of music-induced emotions, cannot be understated, it is essential to consider that the presence of voxels across the entire brain, especially within the three models that decoded the emotional content of music clips, may diminish the significance of specific regions. It is ambitious to claim that one region holds more importance than another in the classification of emotional states.

6.6. Limitations and future work

The present work allowed to successfully achieve the proposed goals: to decode the emotional content of music clips in terms of their valence and arousal levels separately and simultaneously (i.e., to decode each individual quadrant). Nevertheless, there are some factors that should be considered and implemented in the future to allow an improvement of the results and of the generalization ability of decoding models regarding music-evoked emotions.

Foremost, it is relevant to notice that the analyses were performed with spatial smoothing of the fMRI data, and computed in the normalized MNI-space, so significant voxels may have "smeared" from one region into another adjacent region by the smoothing procedure, leading to imprecise interpretations of certain regions being relevant for the decoding tasks.

Additionally, the clusters identified as containing significant information were always very small and dispersed. This was a result of the noisy stability masks used for the

decoding of valence, arousal and of each of the four quadrants individually, that comprised dispersed voxels across the whole brain. Future work may include clustering techniques in the stability masks as a way of condensing the regions that may be used by the classifiers to perform the decoding.

Also, the stability masks used were highly tailored to the neural correlates of each participant while listening to the songs. Whilst this allows for training the models beyond the characteristics of specific songs in inter-subject approach, it might limit the extrapolation power of the classifiers to external groups of subjects. Henceforth, the next step might include the creation of group stability masks in a leave-one-out approach by using the activation levels of all of the participants excepting one and testing on that one.

Furthermore, when the PART labels were used as targets, the small sample size of the testing set and its imbalance in the number of points of each class might be, as hypothesised, the explanation for the discrepancy of results between the two types of labels. Accordingly, an increase of the sample size (i.e., the number of data points of each class) could be considered, though this may lead to a necessity of increase the experimental time, which may cause discomfort to the participants and should be pondered attentively.

Regarding the contributions of voxels in the auditory cortex, further work should include a control condition to account for the contribution of acoustical differences between stimuli and explore the contribution of the musical features of each excerpt to the evoked emotions and classification model performance. This would allow to assess if the obtained results are a consequence of the processing of the emotional content of the music excerpts and not merely of the differences in the auditory processing due to different music features across the four quadrants.

Also, an approach to reduce the possible irrelevant variations between different music excerpts of the same quadrant might include averaging the fMRI images over music clips of the same category (instead of doing solely an average of brain volumes within the same music clip) hence increasing the proportion of signals relevant to emotion categorisation.

Additionally, the discrete classification of the music excerpts in only four classes may be a limiting factor since, even if two music excerpts are situated within the same quadrant, they might have different absolute values of valence and arousal. Therefore, considering the participants ratings in terms of the specific distance to the valence and arousal axis pointed by them when asked to situate the heard excerpts in the Russell's circumplex after the fMRI scanner, instead of solely in terms of the quadrant, may provide a better insight of the subjective perception of the emotional content of the musical clips.

Finally, the assessment of the influence of mood states in the perception and processing of the emotional content of music could be done by integrating the results of the POMS questionnaire answered by the participants and to correlate them with both the behavioural and neural correlates findings.

.

7

Conclusions

The main objective of this work was to define a classification model able to decode the emotional content of a music clip by analysing the fMRI patterns during the listening of different auditory stimuli. Additionally, we aimed to identify the neural correlates of music evoked emotions.

These goals were achieved, with the best decoding results obtained when considering the PRED labels as targets of the classification and stability masks as the feature selection method.

Regarding the optimal model, the most significant regions for the decoding analysis were found in neocortical regions across the whole brain, particularly within the auditory cortex, in the prefrontal cortex, in somatosensory and motor areas as well as across multiple visual areas. Additionally, some clusters were found in less number in some subcortical areas and in the cingulate cortex. These results highlight the role of mechanisms such as rhythmic entrainment, mental imagery and episodic memory as fundamental ways through which music is able to elicit emotions in humans.

The present work provides a valuable framework for the decoding of music evoked emotions, extending previous research by including a larger set of music capable of evoking emotions comprising different categories.

The ability to decode the emotional content of such a broad set of musical *stimuli* in terms of valence and arousal levels supports its promising use to modulate emotional responses and, ultimately, to guide music-based neurofeedback therapies.

Supplementary Material

The supplementary material can be consulted in:
https://drive.google.com/file/d/1KUIB0SbiUuymbQk0kdJeYlcdT30Aofvt/view?usp=sharing

References

[1] I. Cross, "The Nature of Music and Its Evolution," Oct. 2014, doi: 10.1093/OXFORDHB/9780198722946.013.5.

[2] G. F. Welch, M. Biasutti, J. MacRitchie, G. E. McPherson, and E. Himonides, "Editorial: The Impact of Music on Human Development and Well-Being," *Front. Psychol.*, vol. 11, p. 1246, Jun. 2020, doi: 10.3389/FPSYG.2020.01246/BIBTEX.

[3] A. J. Lonsdale and A. C. North, "Why do we listen to music? A uses and gratifications analysis," *Br. J. Psychol.*, vol. 102, no. 1, pp. 108–134, Feb. 2011, doi: 10.1348/000712610X506831.

[4] S. Koelsch, "A coordinate-based meta-analysis of music-evoked emotions," *Neuroimage*, vol. 223, p. 117350, Dec. 2020, doi: 10.1016/J.NEUROIMAGE.2020.117350.

[5] M. L. Chanda and D. J. Levitin, "The neurochemistry of music," *Trends Cogn. Sci.*, vol. 17, no. 4, pp. 179–193, 2013, doi: 10.1016/J.TICS.2013.02.007.

[6] P. N. Juslin, "From everyday emotions to aesthetic emotions: Towards a unified theory of musical emotions," *Phys. Life Rev.*, vol. 10, no. 3, pp. 235–266, Sep. 2013, doi: 10.1016/J.PLREV.2013.05.008.

[7] M. Waterman, "Emotional responses to music: Implicit and explicit effects in listeners and performers," *Psychol. Music*, vol. 24, no. 1, pp. 53–67, 1996, doi: 10.1177/0305735696241006.

[8] I. B. Mauss, L. McCarter, R. W. Levenson, F. H. Wilhelm, and J. J. Gross, "The tie that binds? Coherence among emotion experience, behavior, and physiology," *Emotion*, vol. 5, no. 2, pp. 175–190, Jun. 2005, doi: 10.1037/1528-3542.5.2.175.

[9] J. D. Kropotov, "Affective System," *Quant. EEG, Event-Related Potentials Neurother.*, pp. 292–309, 2009, doi: 10.1016/b978-0-12-374512-5.00013-x.

[10] J. A. Russell, "A circumplex model of affect.," *J. Pers. Soc. Psychol.*, vol. 39, pp. 1161–1178, 1980, doi: 10.1037/h0077714.

[11] 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,

Aug. 2008, doi: 10.1037/1528-3542.8.4.494.

[12] R. E. Thayer, “The biopsychology of mood and arousal,” p. 234, 1989, Accessed: Jun. 13, 2023. [Online]. Available: https://books.google.com/books/about/The_Biopsychology_of_Mood_and_Arousal.html?hl=pt-PT&id=2oe_QgAACAAJ.

[13] S. DiPaola and A. Arya, “Affective communication remapping in musicface system,” *Proc. Eur. Conf. Electron. Imaging Vis. Arts, EVA*, no. May, pp. 1–10, 2004, [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.93.2095&rep=rep1&type=pdf.

[14] K. Hevner, “The Effects of Music, chaps. VII, VIII. 3Max Schoen, The aesthetic attitude in music,” *Psychol. Monog*, vol. 47, no. 2, pp. 162–183, 1935.

[15] R. E. Silva Panda, “AUTOMATIC MOOD TRACKING IN AUDIO MUSIC,” 2010.

[16] J. L. Pool, “Magnetic resonance imaging,” *Biomed. Instrum. Technol.*, vol. 36, no. 5, pp. 341–346, 2002.

[17] J. Sulzer *et al.*, “Real-time fMRI neurofeedback: Progress and challenges,” *Neuroimage*, vol. 76, p. 386, Aug. 2013, doi: 10.1016/J.NEUROIMAGE.2013.03.033.

[18] G. H. Glover, “Overview of Functional Magnetic Resonance Imaging,” *Neurosurg. Clin. N. Am.*, vol. 22, no. 2, p. 133, Apr. 2011, doi: 10.1016/J.NEC.2010.11.001.

[19] R. B. Buxton, “The physics of functional magnetic resonance imaging (fMRI),” *Rep. Prog. Phys.*, vol. 76, no. 9, Sep. 2013, doi: 10.1088/0034-4885/76/9/096601.

[20] N. K. Logothetis and B. A. Wandell, “Interpreting the BOLD signal,” *Annu. Rev. Physiol.*, vol. 66, pp. 735–769, 2004, doi: 10.1146/ANNUREV.PHYSIOL.66.082602.092845.

[21] Y. Yang, W. Engelien, H. Pan, S. Xu, D. A. Silbersweig, and E. Stern, “A CBF-based event-related brain activation paradigm: characterization of impulse-response function and comparison to BOLD,” *Neuroimage*, vol. 12, no. 3, pp. 287–297, 2000, doi: 10.1006/NIMG.2000.0625.

[22] B. Y. Park, K. Byeon, and H. Park, “FuNP (fusion of neuroimaging preprocessing) pipelines: A fully automated preprocessing software for functional magnetic resonance

imaging," *Front. Neuroinform.*, vol. 13, p. 5, Feb. 2019, doi: 10.3389/FNINF.2019.00005/BIBTEX.

[23] M. A. Lindquist, "The Statistical Analysis of fMRI Data," *https://doi.org/10.1214/09-STS282*, vol. 23, no. 4, pp. 439–464, Nov. 2008, doi: 10.1214/09-STS282.

[24] E. T. Bullmore and J. Suckling, "Functional magnetic resonance imaging," *http://dx.doi.org/10.1080/09540260020024169*, vol. 13, no. 1, pp. 24–33, 2009, doi: 10.1080/09540260020024169.

[25] M. M. Monti, "Statistical analysis of fMRI time-series: A critical review of the GLM approach," *Front. Hum. Neurosci.*, vol. 5, no. MARCH, p. 28, Mar. 2011, doi: 10.3389/FNHUM.2011.00028/BIBTEX.

[26] H. Abdulrahman and R. N. Henson, "Effect of trial-to-trial variability on optimal event-related fMRI design: Implications for Beta-series correlation and multi-voxel pattern analysis," *Neuroimage*, vol. 125, p. 756, Jan. 2016, doi: 10.1016/J.NEUROIMAGE.2015.11.009.

[27] 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, Feb. 2012, doi: 10.1016/J.NEUROIMAGE.2011.08.076.

[28] J. D. Haynes, "A Primer on Pattern-Based Approaches to fMRI: Principles, Pitfalls, and Perspectives," *Neuron*, vol. 87, no. 2, pp. 257–270, Jul. 2015, doi: 10.1016/J.NEURON.2015.05.025.

[29] F. Pereira, T. Mitchell, and M. Botvinick, "Machine learning classifiers and fMRI: a tutorial overview.," *Neuroimage*, vol. 45, no. 1 Suppl, 2009, doi: 10.1016/J.NEUROIMAGE.2008.11.007.

[30] M. A. T. Vu *et al.*, "A shared vision for machine learning in neuroscience," in *Journal of Neuroscience*, vol. 38, no. 7, 2018, pp. 1601–1607.

[31] "Overfitting and Underfitting With Machine Learning Algorithms - MachineLearningMastery.com." https://machinelearningmastery.com/overfitting-and-underfitting-with-machine-learning-algorithms/ (accessed Apr. 06, 2023).

[32] "3.1. Cross-validation: evaluating estimator performance — scikit-learn 1.2.2 documentation." https://scikit-learn.org/stable/modules/cross_validation.html (accessed Apr. 12, 2023).

[33] S. Marsland, *Machine Learning: An Algorithmic Perspective, Second Edition*, 2nd ed. Chapman & Hall/CRC, 2014.

[34] "Intro to Feature Selection Methods for Data Science | by Madeline McCombe | Towards Data Science." https://towardsdatascience.com/intro-to-feature-selection-methods-for-data-science-4cae2178a00a (accessed Apr. 06, 2023).

[35] C. Allefeld and J-D Haynes, "Multi-voxel Pattern Analysis," *Brain Mapp. An Encycl. Ref.*, vol. 1, pp. 641–646, Jan. 2015, doi: 10.1016/B978-0-12-397025-1.00345-6.

[36] B. Kijsirikul and N. Ussivakul, "Multiclass support vector machines using adaptive directed acyclic graph," *Proc. 2002 Int. Jt. Conf. Neural Networks. IJCNN'02 (Cat. No.02CH37290)*, vol. 1, pp. 980–985, 2002, doi: 10.1109/IJCNN.2002.1005608.

[37] "Support vector machine - Wikipedia." https://en.wikipedia.org/wiki/Support_vector_machine#Linear_SVM (accessed Apr. 08, 2023).

[38] "Multi-class Classification — One-vs-All & One-vs-One | by Amey Band | Towards Data Science." https://towardsdatascience.com/multi-class-classification-one-vs-all-one-vs-one-94daed32a87b (accessed Apr. 10, 2023).

[39] R. Gholami and N. Fakhari, "Support Vector Machine: Principles, Parameters, and Applications," in *Handbook of Neural Computation*, Academic Press, 2017, pp. 515–535.

[40] "What is the kernel trick? Why is it important? | by Grace Zhang | Medium." https://medium.com/@zxr.nju/what-is-the-kernel-trick-why-is-it-important-98a98db0961d (accessed Jun. 11, 2023).

[41] A. Müller and S. Guido, "Introduction to Machine Learning with Python: A Guide for Data Scientists," 2016.

[42] T. et. all. Hastie, "Springer Series in Statistics The Elements of Statistical Learning," in *The Mathematical Intelligencer*, vol. 27, no. 2, 2009, pp. 417–432.

[43] H. Zhang, "The Optimality of Naive Bayes," Accessed: Apr. 10, 2023. [Online]. Available:

www.aaai.org.

[44] P. N. Juslin and D. Västfjäll, “Emotional responses to music: the need to consider underlying mechanisms,” *Behav. Brain Sci.*, vol. 31, no. 5, 2008, doi: 10.1017/S0140525X08005293.

[45] H. A. Arjmand, J. Hohagen, B. Paton, and N. S. Rickard, “Emotional Responses to Music: Shifts in Frontal Brain Asymmetry Mark Periods of Musical Change,” *Front. Psychol.*, vol. 8, no. DEC, Dec. 2017, doi: 10.3389/FPSYG.2017.02044.

[46] I. Molnar-Szakacs and K. Overy, “Music and mirror neurons: from motion to ’e’motion,” *Soc. Cogn. Affect. Neurosci.*, vol. 1, no. 3, pp. 235–241, 2006, doi: 10.1093/SCAN/NSL029.

[47] S. Davies, “Infectious Music: Music-Listener Emotional Contagion,” *Empathy Philos. Psychol. Perspect.*, Jan. 2012, doi: 10.1093/ACPROF:OSO/9780199539956.003.0010.

[48] T. Eerola, J. K. Vuoskoski, and H. Kautiainen, “Being moved by unfamiliar sad music is associated with high empathy,” *Front. Psychol.*, vol. 7, no. SEP, Sep. 2016, doi: 10.3389/fpsyg.2016.01176.

[49] A. D. Patel, “Language, music, syntax and the brain,” *Nat. Neurosci.*, vol. 6, no. 7, pp. 674–681, Jul. 2003, doi: 10.1038/NN1082.

[50] N. Stogios *et al.*, “Exploring patterns of disturbed eating in psychosis: A scoping review,” *Nutrients*, vol. 12, no. 12, pp. 1–39, Dec. 2020, doi: 10.3390/NU12123883.

[51] R. G. Lewis, E. Florio, D. Punzo, and E. Borrelli, “The Brain’s Reward System in Health and Disease 4.1 The Dopaminergic Mesolimbic System and Reward,” *Robbins and Everitt*, 1988, doi: 10.1007/978-3-030-81147-1_4.

[52] S. N. Haber and B. Knutson, “The Reward Circuit: Linking Primate Anatomy and Human Imaging,” *Neuropsychopharmacology*, vol. 35, pp. 4–26, 2010, doi: 10.1038/npp.2009.129.

[53] J. Sugar and M. B. Moser, “Episodic memory: Neuronal codes for what, where, and when,” *Hippocampus*, vol. 29, no. 12, pp. 1190–1205, Dec. 2019, doi: 10.1002/HIPO.23132.

[54] S. Koelsch, T. Fritz, D. Y. V. Cramon, K. Müller, and A. D. Friederici, “Investigating emotion with music: an fMRI study,” *Hum. Brain Mapp.*, vol. 27, no. 3, pp. 239–250, Mar. 2006, doi: 10.1002/HBM.20180.

[55] A. Lingford-Hughes and N. Kalk, "Clinical neuroanatomy," *Core Psychiatry Third Ed.*, pp. 13–34, Jan. 2012, doi: 10.1016/B978-0-7020-3397-1.00002-1.

[56] O. Shany *et al.*, "Surprise-related activation in the nucleus accumbens interacts with music-induced pleasantness," *Soc. Cogn. Affect. Neurosci.*, pp. 459–470, 2019, doi: 10.1093/scan/nsz019.

[57] K. Mueller *et al.*, "Investigating the dynamics of the brain response to music: A central role of the ventral striatum/nucleus accumbens," 2015, doi: 10.1016/j.neuroimage.2015.05.006.

[58] J. Keller, C. B. Young, E. Kelley, K. Prater, D. J. Levitin, and V. Menon, "Trait anhedonia is associated with reduced reactivity and connectivity of mesolimbic and paralimbic reward pathways," doi: 10.1016/j.jpsychires.2013.05.015.

[59] S. Evers and B. Tölgyesi, "Music and the Cerebellum," *Adv. Exp. Med. Biol.*, vol. 1378, pp. 195–212, 2022, doi: 10.1007/978-3-030-99550-8_13.

[60] D. Kostadinov and M. Häusser, "Reward signals in the cerebellum: Origins, targets, and functional implications," *Neuron*, vol. 110, no. 8, pp. 1290–1303, Apr. 2022, doi: 10.1016/J.NEURON.2022.02.015.

[61] L. M. Parsons, A. Petacchi, J. D. Schmahmann, and J. M. Bower, "Pitch discrimination in cerebellar patients: evidence for a sensory deficit," *Brain Res.*, vol. 1303, pp. 84–96, Nov. 2009, doi: 10.1016/J.BRAINRES.2009.09.052.

[62] C. Lega, T. Vecchi, E. D'Angelo, and Z. Cattaneo, "A TMS investigation on the role of the cerebellum in pitch and timbre discrimination," *Cerebellum & Ataxias*, vol. 3, no. 1, Dec. 2016, doi: 10.1186/S40673-016-0044-4.

[63] T. Hopyan, S. Laughlin, and M. Dennis, "Emotions and their cognitive control in children with cerebellar tumors," *J. Int. Neuropsychol. Soc.*, vol. 16, no. 6, pp. 1027–1038, 2010, doi: 10.1017/S1355617710000974.

[64] B. Tölgyesi and S. Evers, "The impact of cerebellar disorders on musical ability," *J. Neurol. Sci.*, vol. 343, no. 1–2, pp. 76–81, Aug. 2014, doi: 10.1016/J.JNS.2014.05.036.

[65] S. Koelsch, V. K. M. Cheung, S. Jentschke, and J. D. Haynes, "Neocortical substrates of feelings evoked with music in the ACC, insula, and somatosensory cortex," *Sci. Rep.*, vol.

11, no. 1, Dec. 2021, doi: 10.1038/s41598-021-89405-y.

[66] S. Koelsch, S. Skouras, and G. Lohmann, “The auditory cortex hosts network nodes influential for emotion processing: An fMRI study on music-evoked fear and joy,” 2018, doi: 10.1371/journal.pone.0190057.

[67] S. Koelsch, V. K. M. Cheung, S. Jentschke, and J.-D. Haynes, “Neocortical substrates of feelings evoked with music in the ACC, insula, and somatosensory cortex,” *Sci. Reports* /, vol. 11, p. 10119, 123AD, doi: 10.1038/s41598-021-89405-y.

[68] V. Putkinen *et al.*, “Decoding Music-Evoked Emotions in the Auditory and Motor Cortex,” *Cereb. Cortex*, vol. 31, no. 5, pp. 2549–2560, 2021, doi: 10.1093/cercor/bhaa373.

[69] T. Eerola and J. K. Vuoskoski, “A comparison of the discrete and dimensional models of emotion in music,” *Psychol. Music*, vol. 39, no. 1, pp. 18–49, 2011, doi: 10.1177/0305735610362821.

[70] M. E. Sachs, A. Habibi, A. Damasio, and J. T. Kaplan, “Decoding the neural signatures of emotions expressed through sound,” 2018, doi: 10.1016/j.neuroimage.2018.02.058.

[71] J. Kim, S. V. Shinkareva, and D. H. Wedell, “Representations of modality-general valence for videos and music derived from fMRI data,” *Neuroimage*, vol. 148, pp. 42–54, 2017, doi: 10.1016/j.neuroimage.2017.01.002.

[72] F. Zhang, J. P. Wang, J. Kim, T. Parrish, and P. C. M. Wong, “Decoding multiple sound categories in the human temporal cortex using high resolution fMRI,” *PLoS One*, vol. 10, no. 2, pp. 1–19, 2015, doi: 10.1371/journal.pone.0117303.

[73] X. Yang, Y. Dong, and J. Li, “Review of data features-based music emotion recognition methods,” *Multimed. Syst.*, vol. 24, no. 4, pp. 365–389, Jul. 2017, doi: 10.1007/S00530-017-0559-4.

[74] R. Panda, R. Malheiro, and R. P. Paiva, “Novel Audio Features for Music Emotion Recognition,” *IEEE Trans. Affect. Comput.*, vol. 11, no. 4, pp. 614–626, Oct. 2020, doi: 10.1109/TAFFC.2018.2820691.

[75] A. B. Warriner, V. Kuperman, and M. Brysbaert, “Norms of valence, arousal, and dominance for 13,915 English lemmas,” *Behav. Res. Methods*, vol. 45, no. 4, pp. 1191–1207, Dec. 2013, doi: 10.3758/S13428-012-0314-X.

[76] G. Valenza, L. Citi, A. Lanatá, E. P. Scilingo, and R. Barbieri, "Revealing Real-Time Emotional Responses: a Personalized Assessment based on Heartbeat Dynamics," *Sci. Reports 2014 41*, vol. 4, no. 1, pp. 1–13, May 2014, doi: 10.1038/srep04998.

[77] M. S. Soliva *et al.*, "Intervention study to verify the effect of live classic music during hemodialysis on the quality of life of patients with chronic kidney disease," *Nefrologia*, 2023, doi: 10.1016/J.NEFROE.2021.07.010.

[78] K. Takabatake *et al.*, "Musical Auditory Alpha Wave Neurofeedback: Validation and Cognitive Perspectives," *Appl. Psychophysiol. Biofeedback*, vol. 46, no. 4, pp. 323–334, 2021, doi: 10.1007/s10484-021-09507-1.

[79] L. Skottnik, B. Sorger, T. Kamp, D. Linden, and R. Goebel, "Success and failure of controlling the real-time functional magnetic resonance imaging neurofeedback signal are reflected in the striatum," *Brain Behav.*, vol. 9, no. 3, Mar. 2019, doi: 10.1002/BRB3.1240.

[80] B. Direito, M. Ramos, J. Pereira, A. Sayal, T. Sousa, and M. Castelo-Branco, "Directly Exploring the Neural Correlates of Feedback-Related Reward Saliency and Valence During Real-Time fMRI-Based Neurofeedback," *Front. Hum. Neurosci.*, vol. 14, no. February, pp. 1–16, 2021, doi: 10.3389/fnhum.2020.578119.

[81] A. Hillis *et al.*, "Emotion Regulation Using Virtual Environments and Real-Time fMRI Neurofeedback," *Front. Neurol.* / *www.frontiersin.org*, vol. 1, p. 390, 2018, doi: 10.3389/fneur.2018.00390.

[82] M. Brant-Zawadzki, G. D. Gillan, and W. R. Nitz, "MP RAGE: a three-dimensional, T1-weighted, gradient-echo sequence--initial experience in the brain," *Radiology*, vol. 182, no. 3, pp. 769–775, 1992, doi: 10.1148/RADIOLOGY.182.3.1535892.

[83] O. Esteban *et al.*, "fMRIPrep: a robust preprocessing pipeline for functional MRI," *Nat. Methods*, vol. 16, no. 1, pp. 111–116, Jan. 2019, doi: 10.1038/S41592-018-0235-4.

[84] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, Jun. 2002, doi: 10.1613/JAIR.953.

[85] M. A. Casey, "Music of the 7Ts: Predicting and decoding multivoxel fMRI responses with

acoustic, schematic, and categorical music features," *Front. Psychol.*, vol. 8, no. JUL, Jul. 2017, doi: 10.3389/fpsyg.2017.01179.

[86] "Test with permutations the significance of a classification score — scikit-learn 1.2.2 documentation." https://scikit-learn.org/stable/auto_examples/model_selection/plot_permutation_tests_for_classification.html#sphx-glr-auto-examples-model-selection-plot-permutation-tests-for-classification-py (accessed Jun. 29, 2023).

[87] P. Golland and B. Fischl, "LNCS 2732 - Permutation Tests for Classification: Towards Statistical Significance in Image-Based Studies."

[88] P. Mishra, C. M. Pandey, U. Singh, A. Gupta, C. Sahu, and A. Keshri, "Descriptive Statistics and Normality Tests for Statistical Data," *Ann. Card. Anaesth.*, vol. 22, no. 1, p. 67, Jan. 2019, doi: 10.4103/ACA.ACA_157_18.

[89] S. S. Shapiro and M. B. Wilk, "An Analysis of Variance Test for Normality (Complete Samples)," *Biometrika*, vol. 52, no. 3/4, p. 591, Dec. 1965, doi: 10.2307/2333709.

[90] R. Winters, A. Winters, and R. G. Amedee, "Statistics: A Brief Overview."

[91] "Neurosynth: (-56, -20, 12)." https://neurosynth.org/locations/?x=-56&y=-20&z=12 (accessed Sep. 06, 2023).

[92] H. Saarimäki *et al.*, "Discrete Neural Signatures of Basic Emotions," *Cereb. Cortex*, vol. 26, no. 6, pp. 2563–2573, 2016, doi: 10.1093/cercor/bhv086.

[93] L. M. Romanski and J. E. Ledoux, "Information cascade from primary auditory cortex to the amygdala: Corticocortical and corticoamygdaloid projections of temporal cortex in the rat," *Cereb. Cortex*, vol. 3, no. 6, pp. 515–532, 1993, doi: 10.1093/cercor/3.6.515.

[94] L. M. Romanski, J. F. Bates, and P. S. Goldman-Rakic, "Auditory belt and parabelt projections to the prefrontal cortex in the rhesus monkey," *J. Comp. Neurol.*, vol. 403, no. 2, pp. 141–157, 1999, doi: 10.1002/(SICI)1096-9861(19990111)403:2<141::AID-CNE1>3.0.CO;2-V.

[95] M. Yukie, "RESEfl," vol. 0102, no. 95, 1995.

[96] S. Koelsch, V. K. M. Cheung, S. Jentschke, and J. D. Haynes, "Neocortical substrates of feelings evoked with music in the ACC, insula, and somatosensory cortex," *Sci. Rep.*, vol.

11, no. 1, pp. 1–11, 2021, doi: 10.1038/s41598-021-89405-y.

[97] T. Ethofer, D. Van De Ville, K. Scherer, and P. Vuilleumier, "Decoding of Emotional Information in Voice-Sensitive Cortices," *Curr. Biol.*, vol. 19, no. 12, pp. 1028–1033, 2009, doi: 10.1016/j.cub.2009.04.054.

[98] J. Gu, L. Cao, and B. Liu, "Modality-general representations of valences perceived from visual and auditory modalities," *Neuroimage*, vol. 203, no. November 2018, p. 116199, 2019, doi: 10.1016/j.neuroimage.2019.116199.

[99] T. Okuya *et al.*, "Investigating the type and strength of emotion with music: An fMRI study," *Acoust. Sci. Technol.*, vol. 38, no. 3, pp. 120–127, 2017, doi: 10.1250/ast.38.120.

[100] I. R. Olson, A. Plotzker, and Y. Ezzyat, "The Enigmatic temporal pole: A review of findings on social and emotional processing," *Brain*, vol. 130, no. 7, pp. 1718–1731, 2007, doi: 10.1093/brain/awm052.

[101] C. L. Gordon, P. R. Cobb, and R. B. Id, "Recruitment of the motor system during music listening: An ALE meta-analysis of fMRI data," 2018, doi: 10.1371/journal.pone.0207213.

[102] W. Cai, J. S. George, F. Verbruggen, C. D. Chambers, and A. R. Aron, "The role of the right presupplementary motor area in stopping action: two studies with event-related transcranial magnetic stimulation," *J Neurophysiol*, vol. 108, pp. 380–389, 2012, doi: 10.1152/jn.00132.2012.-Rapidly.

[103] M. L. Dixon, R. Thiruchselvam, R. Todd, and K. Christoff, "Emotion and the Prefrontal Cortex : An Integrative Review Challenges in Understanding the Role of the PFC in Emotion," *Psychol. Bull.*, vol. 143, no. 10, pp. 1033–1081, 2017.

[104] M. C. Fasano *et al.*, "The early adolescent brain on music: Analysis of functional dynamics reveals engagement of orbitofrontal cortex reward system," 2022, doi: 10.1002/hbm.26060.

[105] J. Hou *et al.*, "Review on Neural Correlates of Emotion Regulation and Music: Implications for Emotion Dysregulation," *Front. Psychol.*, vol. 8, no. MAR, p. 501, Apr. 2017, doi: 10.3389/FPSYG.2017.00501.

[106] A. D'Argembeau *et al.*, "Distinct regions of the medial prefrontal cortex are associated with self-referential processing and perspective taking," *J. Cogn. Neurosci.*, vol. 19, no.

6, pp. 935–944, Jun. 2007, doi: 10.1162/JOCN.2007.19.6.935.

[107] P. Janata, "The neural architecture of music-evoked autobiographical memories," *Cereb. Cortex*, vol. 19, no. 11, pp. 2579–2594, 2009, doi: 10.1093/CERCOR/BHP008.

[108] K. N. Ochsner *et al.*, "Reflecting upon feelings: an fMRI study of neural systems supporting the attribution of emotion to self and other," *J. Cogn. Neurosci.*, vol. 16, no. 10, pp. 1746–1772, Dec. 2004, doi: 10.1162/0898929042947829.

[109] M. F. Mason, M. I. Norton, J. D. Van Horn, D. M. Wegner, S. T. Grafton, and C. N. Macrae, "Wandering Minds: The Default Network and Stimulus-Independent Thought," Accessed: Aug. 30, 2023. [Online]. Available: www.sciencemag.org/cgi/content/full/[vol]/[issueno.]/[page]/DC1.

[110] W. Trost, T. Ethofer, M. Zentner, and P. Vuilleumier, "Mapping aesthetic musical emotions in the brain," *Cereb. Cortex*, vol. 22, no. 12, pp. 2769–2783, Dec. 2012, doi: 10.1093/CERCOR/BHR353.

[111] S. Koelsch, T. Fritz, K. Schulze, D. Alsop, and G. Schlaug, "Adults and children processing music: An fMRI study," *Neuroimage*, vol. 25, no. 4, pp. 1068–1076, 2005, doi: 10.1016/j.neuroimage.2004.12.050.

[112] V. Nejati, R. Majdi, M. A. Salehinejad, and M. A. Nitsche, "The role of dorsolateral and ventromedial prefrontal cortex in the processing of emotional dimensions," *Sci. Reports 2021 111*, vol. 11, no. 1, pp. 1–12, Jan. 2021, doi: 10.1038/s41598-021-81454-7.

[113] S. Elmer, "Broca pars triangularis constitutes a 'hub' of the language-control network during simultaneous language translation," *Front. Hum. Neurosci.*, vol. 10, no. SEP2016, p. 204888, Sep. 2016, doi: 10.3389/FNHUM.2016.00491/BIBTEX.

[114] H. E. Schaefer, "Music-evoked emotions-Current studies," *Frontiers in Neuroscience*, vol. 11, no. NOV. Frontiers Media S.A., Nov. 24, 2017, doi: 10.3389/fnins.2017.00600.

[115] L. Taruffi, C. Pehrs, S. Skouras, and S. Koelsch, "Effects of Sad and Happy Music on Mind-Wandering and the Default Mode Network," *Sci. Reports 2017 71*, vol. 7, no. 1, pp. 1–10, Oct. 2017, doi: 10.1038/s41598-017-14849-0.

[116] A. E. Cavanna and M. R. Trimble, "The precuneus: A review of its functional anatomy and behavioural correlates," *Brain*, vol. 129, no. 3, pp. 564–583, 2006, doi:

10.1093/brain/awl004.

[117] P. P. Thakral, K. P. Madore, X. Daniel, and L. Schacter, “A Role for the Left Angular Gyrus in Episodic Simulation and Memory,” 2017, doi: 10.1523/JNEUROSCI.1319-17.2017.

[118] H. Platel, J. C. Baron, B. Desgranges, F. Bernard, and F. Eustache, “Semantic and episodic memory of music are subserved by distinct neural networks,” *Neuroimage*, vol. 20, no. 1, pp. 244–256, Sep. 2003, doi: 10.1016/S1053-8119(03)00287-8.