

CENTRO UNIVERSITÁRIO FEI

ESTELA RIBEIRO

**ANÁLISE E RECONHECIMENTO DE PADRÕES COGNITIVOS EM ESCUTAS  
MUSICAIS E SONOROS EM ÁUDIOS**

São Bernardo do Campo

2020

ESTELA RIBEIRO

**ANÁLISE E RECONHECIMENTO DE PADRÕES COGNITIVOS EM ESCUTAS  
MUSICAIS E SONOROS EM ÁUDIOS**

Tese de Doutorado, apresentada ao Centro Universitário FEI para obtenção do título de Doutora em Engenharia Elétrica. Orientada pelo Prof. Dr. Carlos Eduardo Thomaz.

São Bernardo do Campo

2020

Ribeiro, Estela.

Análise e reconhecimento de padrões cognitivos em escutas musicais e sonoros em áudios / Estela Ribeiro. São Bernardo do Campo, 2020.

121 p. : il.

Tese - Centro Universitário FEI.

Orientador: Prof. Dr. Carlos Eduardo Thomaz.

1. Música. 2. EEG. 3. Reconhecimento de padrões. 4. Características acústicas. I. Eduardo Thomaz, Carlos, orient. II. Título.

**Aluno:** Estela Ribeiro

**Matrícula:** 517306-7

**Título do Trabalho:** Análise e reconhecimento de padrões cognitivos em escutas musicais e sonoros.

**Área de Concentração:** Processamento de Sinais e Imagens

**Orientador:** Prof. Dr. Carlos Eduardo Thomaz

**Data da realização da defesa:** 11/12/2020

**ORIGINAL ASSINADA**

Avaliação da Banca Examinadora

A aluna fez a apresentação da sua tese em 45 minutos e respondeu aos questionamentos de cada membro da banca satisfatoriamente. Ao final, foi aprovada por unanimidade. Por sugestão da banca, o título deverá ser alterado para "Análise e reconhecimento de padrões cognitivos em escutas musicais e sonoros em áudios".

São Bernardo do Campo,     /     /     .

**MEMBROS DA BANCA EXAMINADORA**

Prof. Dr. Carlos Eduardo Thomaz	Ass.: _____
Prof. Dr. Ivandro Sanches	Ass.: _____
Profª Drª Patricia Maria Vanzella	Ass.: _____
Prof. Dr. João Ricardo Sato	Ass.: _____
Prof. Dr. Pedro Henrique Alliprandini	Ass.: _____

A Banca Examinadora acima-assinada atribuiu ao aluno o seguinte:

APROVADO

REPROVADO

**VERSÃO FINAL DA TESE**

**ENDOSSO DO ORIENTADOR APÓS A INCLUSÃO DAS  
RECOMENDAÇÕES DA BANCA EXAMINADORA**

Aprovação do Coordenador do Programa de Pós-graduação

\_\_\_\_\_  
Prof. Dr. Carlos Eduardo Thomaz

Dedico esta tese aos meus pais. Sem eles nada seria possível.

## AGRADECIMENTOS

Primeiramente gostaria de agradecer a DEUS pela oportunidade de cursar este doutorado. Eu não teria chegado aonde cheguei se não fosse com a ajuda de Deus e sei que Ele guia cada passo meu. *O Senhor é a minha força e o meu escudo; Nele o meu coração confia, e Dele recebo ajuda.*

Aos meus pais Edivaldo Ribeiro e Dilce Rodrigues da Silva Ribeiro que sempre estiveram ao meu lado me apoiando durante toda a minha trajetória. Vocês sempre me ensinaram a importância de estudar e lutar para alcançar meus objetivos. *Como posso ser suficientemente grata a Deus por vocês, por toda a alegria que tenho diante Dele por causa de vocês?*

A minha irmã Angela Ribeiro de Oliveira pela amizade e atenção dedicadas quando sempre precisei. Você sempre me apoiou e é um exemplo para mim. *O amigo ama em todo momento; é um irmão na adversidade.*

Agradeço ao meu orientador prof. Dr. Carlos Eduardo Thomaz, por aceitar conduzir o meu trabalho de pesquisa e pelas valiosas contribuições dadas durante todo o processo.

Também quero agradecer ao Centro Universitário FEI e a todos os professores do meu curso pela excelência do trabalho de cada um e pela elevada qualidade de ensino oferecido.

Agradeço ao prof. Dr. Ivandro Sanches e a prof. Dr. Patricia Vanzella pelas contribuições durante a banca de qualificação do doutorado. Agradeço também a prof. Patricia pelas reflexões sobre um capítulo que está atualmente em análise que me fizeram redigir de uma forma menos matemático-computacional esta tese, a fim de me aproximar mais do leitor e facilitar a compreensão.

A CAPES e ao Centro Universitário FEI pela bolsa de estudos de doutorado.

Se não houver frutos, valeu a beleza das flores; se não houver flores, valeu a sombra das folhas; se não houver folhas, valeu a intenção da semente. Henfil.

## RESUMO

Estamos envolvidos em um ambiente repleto de sons ao nosso redor. Estudar e analisar os impactos que a prática musical causa e mostrar matematicamente que esta prática pode proporcionar efeitos cognitivos significativos no cérebro humano são as principais motivações desta tese. Em mais detalhes, o objetivo desta tese foi desenvolver uma metodologia capaz de caracterizar os padrões de ativações corticais gerados durante o registro de sinais de Eletroencefalograma (EEG) por meio de técnicas de reconhecimento de padrões em estatística, além de analisar as características acústicas comumente empregadas neste contexto, a fim de revelar se as mesmas são estatisticamente relevantes. Foi desenvolvido inicialmente um arcabouço computacional para abordar o problema de classificação de 2 grupos de amostras baseado em dados de sinais de EEG extraídos de voluntários músicos e não-músicos durante uma tarefa auditiva, para prever se uma determinada pessoa é um músico ou não. Os resultados demonstraram que é possível classificar os grupos amostrados com acurácias que variam de 69.2% a 93.8%, permitindo não somente uma melhor descrição dos padrões de ativações neurais que caracterizam os voluntários músicos e não-músicos, mas também destacando como esses padrões se alteram nas regiões de transição e fronteiras de decisão que separam os grupos amostrados, indicando uma separação linear plausível entre estes grupos. Adicionalmente, como outra contribuição original desta tese, foram analisados os sinais de áudio de uma base de dados pública e internacionalmente referenciada que contém 1000 trechos musicais com 10 gêneros distintos, para investigar similaridades numéricas entre as características acústicas de baixo nível extraídas dos áudios e comumente exploradas na literatura afim. Os resultados obtidos mostram um comportamento de agrupamento similar entre essas características para todas as músicas analisadas, independente do gênero musical. Foi possível então discutir de maneira inédita a relação entre a forma como as características acústicas das músicas são descritas pela literatura e como as mesmas são agrupadas estatisticamente, revelando que a informação que usamos para processar cognitivamente essas características sonoras é implicitamente estatística. Embora todos os métodos descritos e implementados nesta tese sejam baseados em sinais de EEG, acredita-se que os mesmos possam ser estendidos para outros tipos de sinais cognitivos multivariados, como de Imagem de Ressonância Magnética funcional (fMRI), permitindo uma compreensão maior cortical e sub-corticalmente de funcionamento do nosso cérebro durante escutas musicais.

Palavras-chave: Música. EEG. Reconhecimento de padrões. Características acústicas.

## ABSTRACT

We are involved in an environment full of sounds around us. Studying and analyzing the impacts that musical practice causes and showing mathematically that this practice provides significant cognitive effects on the human brain are the main motivations of this thesis. In more detail, the aim of this thesis was to develop a methodology capable of characterizing the cortical activation patterns generated during the register of Electroencephalogram (EEG) signals through pattern recognition techniques in statistics, in addition to analyzing the acoustic features commonly employed in this context, in order to reveal whether they are statistically relevant. A computational framework was initially developed to address a 2 group classification problem based on data from EEG signals extracted from volunteer musicians and non-musicians during an auditory task, to predict whether a particular person is a musician or not. The results showed that it is possible to classify the sampled groups with accuracy ranging from 69.2% to 93.8%, allowing not only a better description of the neural activation patterns that characterize the musician and non-musician volunteers, but also highlighting how these patterns they change in the transition regions and decision boundaries that separate the sampled groups, indicating a plausible linear separation between these groups. Additionally, as another original contribution of this thesis, the audio signals from a public and internationally referenced database containing 1000 musical excerpts with 10 different genres were analyzed to investigate numerical similarities between the short-term acoustic features extracted from the audios and commonly explored in related literature. The results obtained show a similar cluster behavior among these features for all analyzed music, regardless of the musical genre. It was then possible to discuss in an unprecedented way the relationship between the way the acoustic features of songs are described in the literature and how they are grouped statistically, revealing that the information we use to cognitively process these sound features is implicitly statistical. Although all the methods described and implemented in this thesis are based on EEG signals, it is believed that they can be extended to other types of multivariate cognitive signals, such as, for example, functional Magnetic Resonance Imaging (fMRI), allowing a greater cortical and sub-cortical understanding of the functioning of our brain during listening.

Keywords: Music. EEG. Pattern recognition. Acoustic features.

## LISTA DE ILUSTRAÇÕES

Figura 1	–	Composição do encéfalo. . . . .	23
Figura 2	–	Principais lobos do córtex cerebral humano. . . . .	25
Figura 3	–	Divisão do córtex cerebral humano em 52 áreas funcionais distintas de acordo com Korbinian Brodmann. . . . .	26
Figura 4	–	Neurônios do neocórtex arranjados em diferentes camadas utilizando diferentes técnicas de coloração. . . . .	28
Figura 5	–	Estrutura de um neurônio biológico. . . . .	29
Figura 6	–	O potencial de ação. . . . .	30
Figura 7	–	Transmissão sináptica química. . . . .	32
Figura 8	–	Geração do Sinal Elétrico do EEG. . . . .	34
Figura 9	–	Sequência típica dos componentes ERP a partir da apresentação de estímulos auditivos. . . . .	36
Figura 10	–	Ilustração do circuito auditivo. . . . .	41
Figura 11	–	Materiais utilizados para aquisição dos sinais de EEG. . . . .	60
Figura 12	–	Fluxograma das cinco principais etapas do arcabouço computacional proposto para analisar estatisticamente os sinais de EEG, sendo essas: (I) Extração de características acústicas; (II) Seleção de características acústicas; (III) Seleção de Triggers; (IV) Processamento dos sinais de EEG; e (V) Análise estatística multivariada. . . . .	61
Figura 13	–	Descrição das 12 características acústicas utilizadas nesta tese ordenadas de acordo com suas propriedades. . . . .	64
Figura 14	–	Sinal de áudio da música “ <i>The Flight of the Bumblebee</i> ” de Nikolai Rimsky-Korsakov. . . . .	65
Figura 15	–	Sinal da característica acústica RMS extraída da música “ <i>The Flight of the Bumblebee</i> ” de Nikolai Rimsky-Korsakov. . . . .	65
Figura 16	–	Ilustração da ocorrência de <i>triggers</i> para as características Spectral Rolloff e Root Mean Square (RMS) na música Dança Húngara No.5 de Johannes Brahms. . . . .	67
Figura 17	–	Trecho do sinal da característica acústica RMS extraída da música “ <i>The Flight of the Bumblebee</i> ” de Nikolai Rimsky-Korsakov. . . . .	68

Figura 18 – Ilustração do janelamento de um sinal de EEG arbitrário dividido em 6 épocas, delimitadas em 200 ms pré-estímulo e 800 ms pós-estímulo, para 3 tipos de estímulos distintos. . . . .	69
Figura 19 – Carga fatorial das características acústicas extraídas do sinal de áudio, sendo possível observar a formação de <i>clusters</i> entre as seguintes características: Cluster 1 — 1 (RMS), 4 (Spectral Roughness) e 12 (Spectral Flux); Cluster 2 — 8 (Spectral Skewness) e 9 (Spectral Kurtosis); Cluster 3 — 2 (ZCR), 3 (Spectral Rolloff), 5 (Brightness), 6 (Spectral Entropy), 7 (Spectral Flatness), 10 (Spectral Centroid) e 11 (Spectral Spread). . . . .	76
Figura 20 – Espectrogramas de cada característica acústica extraída do sinal de áudio separadas em clusters (linhas verdes tracejadas), apresentando os <i>triggers</i> selecionados (linhas vermelhas) e a característica acústica selecionada (em negrito e sublinhado) para cada cluster. . . . .	78
Figura 21 – Navegação no hiperplano das quatro primeiras componentes principais (PCs), ordenadas da maior para menor variância, da característica acústica <b>Spectral Rolloff</b> gerados pelo PCA no instante 100 ms pós-estímulo: (a) Mapas topográficos de cada PC; (b) Projeção dos dados em cada PC. . . . .	79
Figura 22 – Navegação no hiperplano discriminante da característica acústica <b>Spectral Rolloff</b> gerada pelo LDA no instante 100 ms pós-estímulo: (Cima) Reconstrução do mapa topográfico médio durante a navegação no eixo discriminante entre o grupo de músicos (direita) e não-músicos (esquerda); (Baixo) Projeção dos dados no eixo discriminante. <i>Acurácia: 15.4% (conjunto de teste) e 74.5% (conjunto de treinamento)</i> . . . . .	80
Figura 23 – Navegação no hiperplano das quatro primeiras componentes principais (PCs), ordenadas da maior para menor variância, da característica acústica <b>Root Mean Square (RMS)</b> gerados pelo PCA no instante 100 ms pós-estímulo: (a) Mapas topográficos de cada PC; (b) Projeção dos dados em cada PC. . . . .	81
Figura 24 – Navegação no hiperplano discriminante da característica acústica <b>Root Mean Square (RMS)</b> gerada pelo LDA no instante 100 ms pós-estímulo: (Cima) Reconstrução do mapa topográfico médio durante a navegação no eixo discriminante entre o grupo de músicos (direita) e não-músicos (esquerda); (Baixo) Projeção dos dados no eixo discriminante. <i>Acurácia: 53.8% (conjunto de teste) e 88.0% (conjunto de treinamento)</i> . . . . .	82

Figura 25	– Navegação no hiperplano das quatro primeiras componentes principais (PCs), ordenadas da maior para menor variância, da característica acústica <b>Spectral Kurtosis</b> gerados pelo PCA no instante 100 ms pós-estímulo: (a) Mapas topográficos de cada PC; (b) Projeção dos dados em cada PC. . . . .	83
Figura 26	– Navegação no hiperplano discriminante da característica acústica <b>Spectral Kurtosis</b> gerada pelo LDA no instante 100 ms pós-estímulo: (Cima) Reconstrução do mapa topográfico médio durante a navegação no eixo discriminante entre o grupo de músicos (direita) e não-músicos (esquerda); (Baixo) Projeção dos dados no eixo discriminante. <i>Acurácia: 69.2% (conjunto de teste) e 93.8% (conjunto de treinamento).</i> . . . . .	84
Figura 27	– Carga fatorial das características acústicas extraídas dos 10 diferentes gêneros musicais contidos na base de áudio GTZAN, sendo esses: Blues, Classical, Country, Disco, Hiphop, Jazz, Metal, Pop, Reggae, Rock. . . . .	93
Figura 28	– Distribuição do índice Silhouette variando o número de fatores de 2 a 12 para os 10 gêneros musicais contidos na base de áudio GTZAN. . . . .	93
Figura 29	– Número de $k$ identificados utilizando o índice Silhouette, variando o número de fatores de 2 a 12 para os 10 gêneros musicais contidos na base de áudio GTZAN. . . . .	94
Figura 30	– Distribuição do índice Davis-Bouldin variando o número de fatores de 2 a 12 para os 10 gêneros musicais contidos na base de áudio GTZAN. . . . .	94
Figura 31	– Número de $k$ identificados utilizando o índice Davis-Bouldin, variando o número de fatores de 2 a 12 para os 10 gêneros musicais contidos na base de áudio GTZAN. . . . .	95
Figura 32	– Matriz de correlação do Fator 1 entre os 10 gêneros musicais contidos na base de áudio GTZAN, utilizando 2 fatores extraídos da FA. . . . .	96
Figura 33	– Matriz de correlação do Fator 2 entre os 10 gêneros musicais contidos na base de áudio GTZAN, utilizando 2 fatores extraídos da FA. . . . .	97
Figura 34	– Protótipo das cargas fatoriais das características acústicas extraídas de todos os áudios contidos na base de áudio GTZAN. . . . .	98
Figura 35	– Relação entre os 12 descritores sonoros de baixo nível utilizados neste trabalho e os 3 clusters encontrados pela FA. . . . .	100

## LISTA DE TABELAS

Tabela 1	– Descrição dos voluntários que participaram dos experimentos. . . . .	75
Tabela 2	– Carga fatorial das características acústicas extraídas do sinal de áudio. . . .	77
Tabela 3	– Taxa de classificação do LDA para todos os eletrodos, 6 eletrodos, 4 eletrodos e 2 eletrodos. . . . .	82
Tabela 4	– Taxa de classificação detalhada do LDA utilizando todos os eletrodos. . . .	84
Tabela 5	– Carga fatorial das variáveis para o gênero <i>Blues</i> . . . . .	119
Tabela 6	– Carga fatorial das variáveis para o gênero <i>Classical</i> . . . . .	119
Tabela 7	– Carga fatorial das variáveis para o gênero <i>Country</i> . . . . .	119
Tabela 8	– Carga fatorial das variáveis para o gênero <i>Disco</i> . . . . .	119
Tabela 9	– Carga fatorial das variáveis para o gênero <i>Hiphop</i> . . . . .	120
Tabela 10	– Carga fatorial das variáveis para o gênero <i>Jazz</i> . . . . .	120
Tabela 11	– Carga fatorial das variáveis para o gênero <i>Metal</i> . . . . .	120
Tabela 12	– Carga fatorial das variáveis para o gênero <i>Pop</i> . . . . .	120
Tabela 13	– Carga fatorial das variáveis para o gênero <i>Reggae</i> . . . . .	120
Tabela 14	– Carga fatorial das variáveis para o gênero <i>Rock</i> . . . . .	120

## LISTA DE ABREVIATURAS

A1	Córtex auditivo primário
DB	Davies-Boulding
EEG	Eletroencefalograma
ERP	Potencial Relacionado ao Evento (do inglês, <i>Event Related Potential</i> )
FA	Análise Fatorial (do inglês, <i>Factor Analysis</i> )
FFT	Transformada rápida de Fourier (do inglês, <i>Fast Fourier Transform</i> )
fMRI	Imagem de Ressonância Magnética funcional (do inglês, <i>functional Magnetic Resonance Imaging</i> )
fNIRS	Espectroscopia funcional em Infravermelho Próximo (do inglês, <i>Functional Near-Infrared Spectroscopy</i> )
LDA	Análise Linear Discriminante (do inglês, <i>Linear Discriminant Analysis</i> )
M1	Córtex motor primário
MEG	Magnetoencefalografia (do inglês, <i>Magnetoencephalography</i> )
MIR	Recuperação de Informação Musical (do inglês, <i>Music Information Retrieval</i> )
MLDA	<i>Maximum uncertainty Linear Discriminant Analysis</i>
MMN	<i>Mismatch negativity</i>
MRI	Imagem por Ressonância Magnética (do inglês, <i>Magnetic Resonance Imaging</i> )
PA	Potencial de Ação
PC	Componente Principal
PCA	Análise de Componentes Principais (do inglês, <i>Principal Component Analysis</i> )
PEPS	Potencial excitatório pós-sináptico
PET	Tomografia por Emissão de Pósitrons (do inglês, <i>Positron Emission Tomography</i> )
PIPS	Potencial inibitório pós-sináptico
PLFP	<i>Preceding Low Feature Phase</i>
RMS	<i>Root Mean Square</i>
S1	Córtex somatossensorial primário
Si	Silhouette
SVM	Máquina de Vetores de Suporte (do inglês, <i>Support Vector Machine</i> )
ZCR	<i>Zero Crossing Rate</i>

## LISTA DE SÍMBOLOS

$\mathbf{A}_g$	Matriz de dados composta pelos valores extraídos das $s$ características acústicas de cada $g$ gênero musical representado na base de dados GTZAN
$b_i$	Distância média mínima entre o ponto $i$ e os pontos nos diferentes agrupamentos
$C_k$	Agrupamentos em $k$ clusters
$c$	Número de classes, $c = 2$
$DB$	Critério Davies-Bouldin
$D_i$	Distância intra-cluster e inter-cluster
$\varepsilon$	Vetor de erros
$\mathbf{F}$	Vetor de fatores comuns
$g$	Gêneros musicais
$H$	Total de músicas contidas na base de dados GTZAN
$h$	$h \in \{-3, -2, -1, 0, 1, 2, 3\}$
$i$	Grupo de amostragem correspondente
$J$	Número de janelas extraídas do sinal de áudio
$j$	$j \in \{-1, 0, 1\}$
$k$	$k$ cluster do algoritmo $k$ -means
$\mathbf{L}$	Matriz de carga fatorial
$\lambda_{m,r}$	Autovalores das $m = 1, 2, \dots, M$ componentes principais para cada característica acústica $r$ selecionada
$M$	Número de componentes principais extraídas de $\mathbf{X}_r$
$m$	Índice da componente principal correspondente
$\mu$	Vetor médio
$N$	Número total de voluntários (sujeitos)
$n$	Número total de eletrodos, $n = 16$
$o_i$	Distância média do ponto $i$ aos outros pontos dentro de $C_o$
$\Omega_d$	Conjunto de treinamento
$\Omega_t$	Conjunto de teste
$\mathbf{p}_{m,r}$	Autovetores das $m = 1, 2, \dots, M$ componentes principais para cada característica acústica $r$ selecionada
$q$	Porcentagem para extração de triggers

$R$	Fatores extraídos da Análise Fatorial (FA), $R \leq s$
$r$	$r = 1, 2, \dots, sf$
$s$	Características acústicas extraídas dos áudios, $s = 12$
$S_b$	Matriz inter-classe do LDA
$sf$	Número de características acústicas selecionadas
$Si$	Critério Silhouette
$\sigma_i$	Desvio padrão de cada grupo amostrado no espaço do LDA para cada característica acústica $r$ selecionada
$S_w$	Matriz intra-classe do LDA
$t$	Instante de análise do sinal de EEG (pré-definido), $t = 100$ ms pós-estímulo
$V_{p-}$	Limiar inferior para extração de triggers
$V_{p+}$	Limiar superior para extração de triggers
$\mathbf{w}_r$	Hiperplano de projeção do LDA (eixo de discriminância)
$\mathbf{X}_r$	Matriz de dados composta pelos valores das épocas médias do sinal de EEG no instante $t = 100$ ms pós-estímulo para cada característica acústica $r$ selecionada
$\bar{x}_i$	Média de cada grupo amostrado no espaço do LDA para cada característica acústica $r$ selecionada
$\bar{\mathbf{x}}_r$	Vetor de média global $n$ -dimensional de $\mathbf{X}_r$ para cada característica acústica $r$ selecionada
$\mathbf{y}_{m,r}$	Hiperplano de navegação do PCA
$\mathbf{z}_{i,j}$	Hiperplano de navegação do LDA

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	18
1.1	MOTIVAÇÃO	20
1.2	OBJETIVO	21
1.3	PUBLICAÇÕES	21
1.4	ORGANIZAÇÃO DO TRABALHO	22
<b>2</b>	<b>GERAÇÃO E REGISTRO DOS SINAIS CEREBRAIS</b>	23
2.1	O SISTEMA NERVOSO E OS SINAIS CEREBRAIS	23
<b>2.1.1</b>	<b>Propriedades das células nervosas</b>	27
2.2	GERAÇÃO E DINÂMICA DOS SINAIS DE EEG	33
<b>2.2.1</b>	<b>Potencial relacionado ao evento</b>	35
2.3	CONSIDERAÇÕES COMPLEMENTARES	38
<b>3</b>	<b>A MÚSICA E O CÉREBRO</b>	39
3.1	COMO O CÉREBRO PROCESSA A MÚSICA	40
3.2	DIFERENÇAS COGNITIVAS ENTRE MÚSICOS E NÃO-MÚSICOS	44
3.3	A MÚSICA E O ENVELHECIMENTO	50
3.4	RESERVA COGNITIVA	53
3.5	DESCRITORES SONOROS	55
3.6	CONSIDERAÇÕES COMPLEMENTARES	58
<b>4</b>	<b>PADRÕES COGNITIVOS EM ESCUTAS MUSICAIS</b>	59
4.1	MATERIAIS	59
4.2	METODOLOGIA	60
<b>4.2.1</b>	<b>Extração de características acústicas</b>	63
<b>4.2.2</b>	<b>Seleção das características acústicas</b>	66
<b>4.2.3</b>	<b>Seleção de triggers</b>	66
<b>4.2.4</b>	<b>Processamento dos sinais de EEG</b>	69
<b>4.2.5</b>	<b>Análise estatística multivariada</b>	70
<b>4.2.5.1</b>	<i>Análise de componentes principais (PCA)</i>	71
<b>4.2.5.2</b>	<i>Análise linear discriminante (LDA)</i>	72
<b>4.2.5.3</b>	<i>Classificação</i>	73
4.3	RESULTADOS	74
<b>4.3.1</b>	<b>Participantes</b>	74

4.3.2	<b>Estímulo</b>	74
4.3.3	<b>Análise do sinal de áudio</b>	75
4.3.4	<b>Análise do sinal de EEG</b>	77
4.4	<b>DISCUSSÃO</b>	82
5	<b>PADRÕES SONOROS EM ÁUDIOS MUSICAIS</b>	88
5.1	<b>MATERIAIS E MÉTODOS</b>	88
5.1.1	<b>Base de áudio</b>	88
5.1.2	<b>Metodologia</b>	89
5.1.2.1	<i>Silhouette (Si)</i>	91
5.1.2.2	<i>Davies-Bouldin (DB)</i>	91
5.2	<b>RESULTADOS</b>	92
5.3	<b>DISCUSSÃO</b>	95
5.4	<b>CONSIDERAÇÕES COMPLEMENTARES: PADRÕES COGNITIVOS E SONOROS</b>	99
6	<b>CONCLUSÃO</b>	102
	<b>REFERÊNCIAS</b>	104
	<b>APÊNDICE A – Descrição das características acústicas</b>	113
	<b>APÊNDICE B – Tabelas</b>	118

## 1 INTRODUÇÃO

O cérebro humano tem uma grande capacidade de se auto-organizar em resposta a estímulos externos. A música oferece uma oportunidade única para compreender esse processo de organização neural (PERETZ; ZATORRE, 2005). Investigações sobre os mecanismos envolvidos em ouvir e fazer música têm contribuído para o entendimento de como o cérebro se altera em resposta ao treinamento e à aquisição de novas habilidades (HUTCHINSON et al., 2003; CHEN; PENHUNE; ZATORRE, 2008).

Tocar um instrumento musical é uma tarefa que requer uma grande demanda neural (VUUST et al., 2012; ANGULO-PERKINS et al., 2014). Imagine que para um músico executar o seu instrumento, ele precisa reter e manipular as informações que foram aprendidas e atualizá-las continuamente, fazendo uso da memória de longo prazo e também da memória de trabalho. Além disso, necessita controlar movimentos complexos e monitorar o ambiente, integrando informações visuais, auditivas e somestésicas<sup>1</sup>, para realizar os ajustes necessários durante a performance, havendo uma forte ligação entre a percepção do som produzido e a execução de movimentos (BROWN; ZATORRE; PENHUNE, 2015; ALTENMÜLLER; SCHLAUG, 2015). Assim, embora o ser humano possua uma habilidade musical implícita, sendo capaz de ouvir e compreender a música naturalmente (KOELSCH et al., 2000), é a prática musical que confere às pessoas alterações a longo prazo que resultam do processo de plasticidade cerebral<sup>2</sup> (MÜNTE; ALTENMÜLLER; JÄNCKE, 2002; ALTENMÜLLER; SCHLAUG, 2015) podendo ser somada, ou não, a fatores inatos (BARRETT et al., 2013).

Estudos têm demonstrado que a prática musical pode proporcionar diversos benefícios tanto cognitivos como terapêuticos, tais como proteção contra o declínio cognitivo relacionado ao envelhecimento (EGGERMONT, 2014; KOK, 2000; SARKAMO, 2017), melhoras em habilidades linguísticas (SARKAMO, 2017; CARPENTIER; MORENO; MCINTOSH, 2016; BIDELMAN; ALAIN, 2015), regulação do humor, evocação de memórias, manutenção da autoestima, redução de sentimentos de solidão e isolamento, entre outros (SARKAMO, 2017; ALTENMÜLLER; SCHLAUG, 2015). Em poucas palavras, a música pode contribuir de várias formas ao bem-estar e à qualidade de vida, com potencial para ser aplicada, por exemplo, na

<sup>1</sup>Somestesia está relacionada com a percepção do corpo por meio do tato, dos movimentos corporais, da posição dos membros no espaço, da temperatura e da dor (LENT, 2001).

<sup>2</sup>Plasticidade cerebral se refere a capacidade do sistema nervoso de alterar a sua estrutura e funcionamento ao longo da vida de acordo com as necessidades e fatores do meio ambiente que está contido (LENT, 2001).

reabilitação e auxílio durante o processo de envelhecimento (BAIRD; SAMSON, 2015; SARKAMO, 2017).

A música, na verdade, é capaz de promover mudanças profundas, tanto estruturais quanto funcionais, em nosso sistema nervoso, e pesquisas evidenciam as diferenças existentes entre o cérebro de músicos e não-músicos (MÜNTE; ALTENMÜLLER; JÄNCKE, 2002; TERVANIEMI et al., 2006). Essas mudanças podem ser observadas em várias áreas encefálicas. Nota-se, por exemplo, que músicos apresentam um aumento na conexão entre os dois hemisférios cerebrais, quando comparados a não-músicos (SCHLAUG et al., 1995), assim como em áreas auditivas, áreas de integração multissensorial e áreas somatossensoriais (SCHNEIDER et al., 2002; SLUMING et al., 2002; GASER; SCHLAUG, 2003b,a; BANGERT; SCHLAUG, 2006; ELMER; HANGGI; JANCKE, 2016). Além disso, a experiência musical ao longo da vida produz impactos significativos nas funções cognitivas. No entanto, diferentes tipos de atividades podem demandar o refinamento de habilidades específicas. Alguns instrumentista, por exemplo, necessitam da coordenação precisa de movimentos bimanuais (um pianista pode chegar a ter que coordenar a produção de cerca de 1800 notas por minuto), enquanto um maestro, mais do que qualquer outro músico, depende muito em sua prática de uma aprimorada habilidade de localização espacial auditiva (MÜNTE; ALTENMÜLLER; JÄNCKE, 2002). Ou seja, mesmo dentro da classe “músicos”, diferentes atividades podem implicar diferentes demandas neurais.

Por meio do uso de ferramentas como o Eletroencefalograma (EEG), a Imagem por Ressonância Magnética Funcional (fMRI), entre outras, um número crescente de pesquisas tem buscado compreender como o cérebro processa a música (MÜNTE; ALTENMÜLLER; JÄNCKE, 2002; PERETZ; ZATORRE, 2005; TERVANIEMI et al., 2006; VUUST et al., 2012; MIKUTTA et al., 2014; VIRTALA et al., 2014; SCHLAUG et al., 2005; SAARI et al., 2018). Grande parte destes estudos tem utilizado estímulos artificiais, construídos especificamente para os experimentos, tais como diferentes sequências sonoras, muitas vezes sintetizadas, tons puros e tons complexos, acordes maiores e menores, acordes consonantes e dissonantes, etc. No entanto, estes estímulos representam apenas algumas das múltiplas dimensões que constituem o que chamamos de música e, portanto, o que esses estudos na realidade mostram é como o cérebro processa estes elementos isoladamente. Estudos mais recentes têm buscado contornar esse problema utilizando como estímulos gravações de músicas completas (ALLURI et al., 2012, 2013; POIKONEN et al., 2016; SAARI et al., 2018; RIBEIRO; THOMAZ, 2019), com a finalidade de descrever a associação entre as mudanças dinâmicas dos áudios com as ativa-

ções neurais registradas nos voluntários. Analisando padrões de ativações neurais gerados em resposta à escuta musical, alguns desses estudos demonstraram que é possível indicar se uma pessoa recebeu, ou não, treinamento musical em sua vida (LIANG et al., 2011; SAARI et al., 2018; RIBEIRO; THOMAZ, 2019). Para tanto, é necessário que os dados sejam representados adequadamente, levando em consideração o tipo de estímulo utilizado durante a aquisição de sinais (sejam esses de fMRI, EEG ou outros), e que sejam utilizadas ferramentas de reconhecimento de padrões e aprendizado de máquina para realizar a classificação dessas informações.

Esta tese propõe, descreve e implementa uma metodologia para identificar diferenças neurais existentes entre músicos e não-músicos por meio de uma análise de dados de EEG registrados em voluntários durante uma tarefa de escuta musical. Embora músicos possam diferir entre si em termos de processamento (devido ao fato de que atividades distintas possam recrutar diferentes recursos neurais), avalia-se aqui se durante a escuta musical é possível separar o grupo de músicos do de não-músicos (RIBEIRO; THOMAZ, 2019). Essa abordagem matemático-computacional proposta permite analisar os sinais de EEG global e holisticamente, permitindo uma melhor descrição dos padrões de ativações neurais que caracterizam os voluntários músicos e não-músicos, destacando como esses padrões se alteram nas regiões de transição e fronteiras de decisão que separam os grupos amostrados, indicando uma separação linear plausível entre estes grupos.

Adicionalmente, como outra contribuição original desta tese, são analisados os sinais de áudio de uma base de dados pública e internacionalmente referenciada, revelando se todas as características acústicas de baixo nível extraídas dos áudios, e comumente explorados na literatura afim, são estatisticamente relevantes. Os resultados obtidos mostram um comportamento similar de agrupamento (*cluster*) entre essas características para as diferentes músicas, independente do gênero musical (FERREIRA; RIBEIRO; THOMAZ, 2019; RIBEIRO; THOMAZ, 2019).

## 1.1 MOTIVAÇÃO

Estamos envolvidos em um ambiente repleto de sons ao nosso redor e a música possui um papel importante em nossas vidas. Portanto, estudar e analisar os impactos que a prática musical causa e mostrar matematicamente que esta prática proporciona efeitos cognitivos significativos no cérebro humano são as principais motivações desta tese.

## 1.2 OBJETIVO

O objetivo desta tese é desenvolver uma metodologia que seja capaz de caracterizar os padrões de ativações corticais gerados durante o registro de sinais de EEG por meio de técnicas de reconhecimento de padrões em estatística, podendo ser escalável para diversos problemas similares de dados cognitivos multidimensionais. Mais especificamente, pretende-se:

- a) Caracterizar os padrões de ativação corticais apresentados por voluntários saudáveis durante uma tarefa auditiva;
- b) Discriminar os grupos de voluntários (músicos e não-músicos) por meio de ferramentas de reconhecimento de padrões em estatística;
- c) Analisar as características acústicas comumente utilizadas a fim de revelar se as mesmas são estatisticamente relevantes ou não.

## 1.3 PUBLICAÇÕES

A maior parte da metodologia, dos experimentos e dos resultados desta tese foram publicados como artigos nas seguintes contribuições científicas:

- a) RIBEIRO, Estela; THOMAZ, Carlos Eduardo. A Whole Brain EEG Analysis of Musicianship. *MUSIC PERCEPTION*, v. 37, p. 42-56, 2019;
- b) FERREIRA, L. A.; RIBEIRO, Estela; THOMAZ, C. E. A cluster analysis of benchmark acoustic features on Brazilian music. In: 17th Brazilian Symposium on Computer Music, 2019, São João del Rei. *SBCM 2019*, 2019. v. 17.
- c) RIBEIRO, E.; THOMAZ, C. E. . A multivariate statistical analysis of EEG signals for differentiation of musicians and non-musicians. In: XV Encontro Nacional de Inteligência Artificial e Computacional, 2018, São Paulo. *ENIAC*, 2018;
- d) RIBEIRO, E.; THOMAZ, C. E. . Um estudo sobre predição de musicalidade por meio da análise de sinais de EEG. In: VIII Simpósio de Instrumentação e Imagens Médicas (SIIM) / VII Simpósio de Processamento de Sinais, 2017, São Bernardo do Campo. *Anais do SIIM / SPS 2017*, 2017.

## 1.4 ORGANIZAÇÃO DO TRABALHO

Este trabalho está organizado da seguinte forma. No Capítulo 2, são apresentados os conceitos fundamentais para a compreensão desta tese. Neste capítulo é descrito, de modo geral, o funcionamento do sistema nervoso e a origem dos sinais cerebrais, assim como o funcionamento do equipamento de eletroencefalografia e o potencial relacionado ao evento, uma técnica tradicionalmente utilizada para analisar os sinais cerebrais. O Capítulo 3 discute a relação entre a música e o cérebro, por meio de uma revisão da literatura, descrevendo como o cérebro processa a música, quais são as diferenças estruturais e funcionais entre músicos e não-músicos, além de abordar brevemente sobre o impacto da música no envelhecimento, apresentando alguns conceitos de reserva cognitiva. Neste capítulo são também apresentadas as características acústicas, que descrevem o som quantitativa ou qualitativamente, utilizadas para se analisar os sinais de áudio empregados nesta tese. O Capítulo 4 propõe um arcabouço computacional para analisar globalmente dados de EEG registrados durante uma tarefa auditiva, por meio de uma análise estatística multivariada, a fim de discriminar músicos e não-músicos. Os resultados deste capítulo demonstram que é possível diferenciar músicos e não-músicos linearmente, apesar do número limitado de amostras. Estes resultados sugerem ainda que é possível evidenciar as variações mais expressivas e discriminantes dos padrões de atividades cerebrais dos participantes, por meio do método de navegação no hiperplano, dependentes da característica acústica extraída do áudio. O Capítulo 5 propõe a identificação apropriada de descritores dos sinais de áudio com o objetivo de gerar uma melhor representação dos estímulos auditivos usados para investigar como o cérebro processa a música. Para isso, é utilizada outra técnica de estatística multivariada com o intuito de revelar se todas as características acústicas de baixo nível extraídas de uma base de dados internacionalmente referenciada são estatisticamente relevantes, ou seja, não-redundantes. Os resultados deste capítulo mostram que nem todas as características acústicas extraídas são necessárias, demonstrando experimentalmente que há redundâncias entre esse conjunto de descritores que não estão relacionados particularmente a uma música ou a um determinado gênero musical. Por fim, o Capítulo 6 discute os resultados obtidos e apresenta propostas para trabalhos futuros.

## 2 GERAÇÃO E REGISTRO DOS SINAIS CEREBRAIS

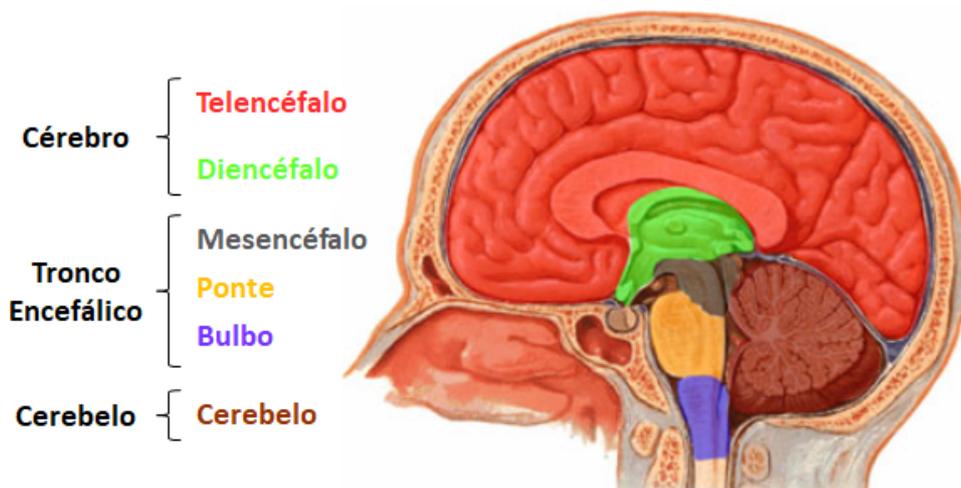
Este capítulo apresenta os conceitos fundamentais para a compreensão deste trabalho. Descreve-se basicamente como o sistema nervoso é constituído, cujas estruturas são de extrema importância para o funcionamento do organismo, assim como a estrutura dos neurônios, que são células altamente especializadas e essenciais para o sistema nervoso, capazes de transmitir informações por meio de sequências de pulsos elétricos chamados Potenciais de Ação e capazes de se comunicarem entre si por meio da Transmissão Sináptica. É também apresentado o Eletroencefalograma, revelando como o sinal elétrico é gerado no cérebro e as suas características, descrevendo também o Potencial Relacionado ao Evento, que é uma técnica utilizada para o registo de potenciais elétricos evocados a partir de um determinado estímulo.

### 2.1 O SISTEMA NERVOSO E OS SINAIS CEREBRAIS

O sistema nervoso possui um papel fundamental para a vida, permitindo ao ser humano interagir com o ambiente, sentir, pensar e se mover. O encéfalo é responsável por diversas funções como pensamento, linguagem, aprendizado, memória, imaginação, criatividade, atenção, consciência, emoção e sono (KREBS; WEIBERG; AKESSON, 2011).

O Encéfalo pode ser dividido em: telencéfalo, diencefalo, tronco encefálico e cerebelo (WECKER; SOARES; NEMOS, 2001). A Figura 1 apresenta a distribuição dessas estruturas que serão brevemente descritas a seguir.

Figura 1 – Composição do encéfalo.



Fonte: Adaptado de Wecker, Soares e Nemos, 2001.

O tronco encefálico fornece o meio pelo qual toda a informação (ascendente e descendente) entre o cérebro e a coluna espinhal trafega. Aspectos das funções cardiovasculares e respiratórias, atividade cortical e consciência são organizadas e integradas em diferentes áreas do tronco encefálico, que é formado pelo Bulbo, a Ponte e o Mesencéfalo (KREBS; WEIBERG; AKESSON, 2011). O Bulbo apresenta centros responsáveis por funções autônomas vitais como a digestão, respiração e controle da frequência cardíaca, além do centro vasomotor e o centro do vômito. A Ponte transmite informações sobre movimento dos hemisférios para o cerebelo, além de ter papel fundamental na regulação do padrão e ritmo respiratório. O Mesencéfalo controla várias funções motoras e sensoriais, incluindo o movimento dos olhos e a coordenação dos reflexos auditivos e visuais (KANDEL et al., 2012).

A parte de composição do encéfalo denominado cerebelo é um centro para controle do movimento, gerando sinais de correção que auxiliam para tornar os movimentos mais precisos. Essa estrutura cerebral atua modulando a força, o alcance, o equilíbrio e a coordenação, e está envolvido com o aprendizado de habilidades motoras (BEAR; CONNORS; PARADISO, 2015).

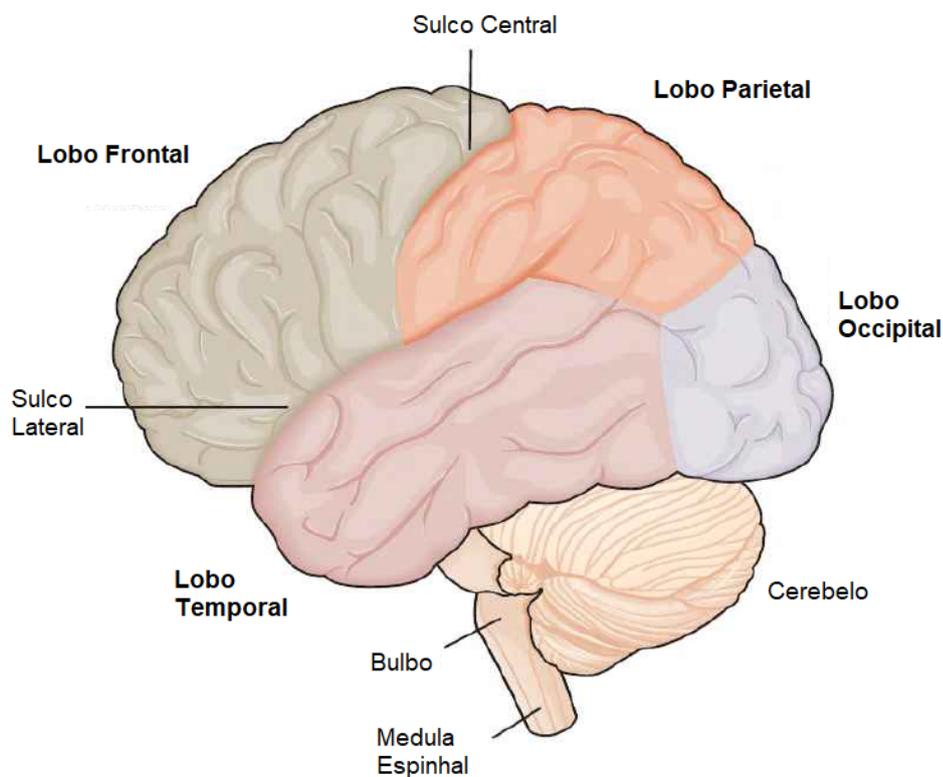
O Diencefalo, por sua vez, contém duas estruturas: O Tálamo e o Hipotálamo. O Tálamo processa maior parte da informação sensorial que chega ao córtex cerebral, apresentando um papel importante no processamento da informação motora, além da integração do comportamento emocional. O Hipotálamo regula a homeostase e diversos comportamentos reprodutivos, além de funções endócrinas e viscerais (KANDEL et al., 2012).

Adicionalmente, os Hemisférios Cerebrais possuem: o Telencéfalo, os Núcleos da Base, a Amígdala, as Formações Hipocâmpais e o Corpo Caloso. O Telencéfalo possui funções cognitivas, perceptuais e motoras, além da memória e emoção. É dividido em hemisférios direito e esquerdo, que são separados pela fissura longitudinal. Os dois hemisférios são interligados pelo Corpo Caloso, que une áreas simétricas do córtex de cada hemisfério. Sabe-se que o treinamento musical promove um aumento na interconexão entre os dois hemisférios, ou seja, um aumento no corpo caloso, por meio da plasticidade cerebral, como resultado da experiência adquirida ao longo de sua prática (SCHLAUG et al., 1995; MÜNTE; ALTENMÜLLER; JÄNCKE, 2002). A Amígdala desempenha um papel relevante no processamento de emoções e formação de memórias, o Hipocampo também está envolvido na regulação da emoção, motivação, além do aprendizado e memória, e os Núcleos da Base com o controle do movimento e aspectos do aprendizado motor (KANDEL et al., 2012).

Sabe-se que a Medula Espinhal, o Tronco Encefálico e o Diencefalo são responsáveis por muitas funções de sustentação da vida, mas as funções de planejamento e execução de

ações ocorrem no Córtex Cerebral (uma camada fina e mais externa dos hemisférios cerebrais) (BEAR; CONNORS; PARADISO, 2015). É no córtex cerebral que ocorrem as principais operações responsáveis pela capacidade cognitiva humana como, por exemplo, o pensamento consciente, a percepção, o aprendizado associativo e o controle executivo sobre o movimento voluntário (SQUIRE et al., 2008). O córtex cerebral é dividido em quatro lobos denominados em função dos ossos cranianos que se situam sobre os próprios lobos, conforme ilustrado na Figura 2 (KANDEL et al., 2012; KREBS; WEIBERG; AKESSON, 2011), sendo:

Figura 2 – Principais lobos do córtex cerebral humano.

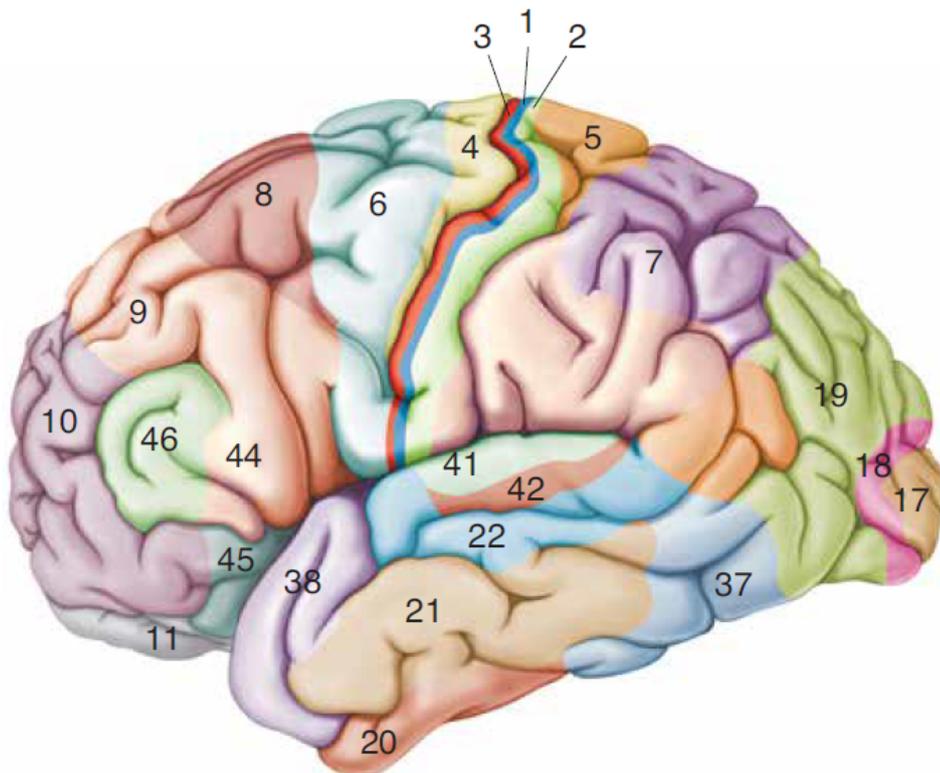


Fonte: Adaptado de Kandel et al., 2012.

- a) Lobo Frontal, envolvido com a memória de curto prazo, planejamento de ações futuras, controle do movimento;
- b) Lobo Parietal, envolvido com a sensação somática, formação de uma imagem corporal e sua relação com o espaço extrapessoal;
- c) Lobo Occipital, envolvido com a visão;
- d) Lobo Temporal, envolvido com a audição, além do aprendizado, memória e emoção por meio do hipocampo e da amígdala que são suas estruturas profundas.

Essa separação em lobos serve como referência para a organização funcional do cérebro (BEAR; CONNORS; PARADISO, 2015). Já foram identificadas mais de 50 áreas que apresentam características estruturais e funcionais distintas, sendo o mapa de áreas corticais desenvolvido por Korbinian Brodmann, no início do século XX, o mais conhecido (SQUIRE et al., 2008). A Figura 3 ilustra este mapa, sendo que diversas áreas definidas por Brodmann correspondem a funções encefálicas específicas, como, por exemplo, a área 4 (córtex motor) relacionado com o movimento voluntário, a área 17 (córtex visual primário) que recebe os sinais sensoriais dos olhos, ou as áreas 41 e 42 que constituem o córtex auditivo primário (KANDEL et al., 2012).

Figura 3 – Divisão do córtex cerebral humano em 52 áreas funcionais distintas de acordo com Korbinian Brodmann.



Fonte: Adaptado de Bear, Connors e Paradiso, 2015.

Uma característica da organização do córtex cerebral é que cada hemisfério está relacionado com processos sensoriais e motores contralaterais (lado oposto) ao corpo. Diferentes áreas do córtex são interconectadas por vias associativas complexas e os hemisférios cerebrais não são completamente simétricos em estrutura nem equivalentes em função (SQUIRE et al., 2008).

Além disso, a região mais superficial do córtex (neocórtex) é organizada em camadas e colunas<sup>1,2</sup> (Figura 4), apresentando diferentes fluxos de informação, variando a sua organização funcional (KANDEL et al., 2012). Essa estrutura laminar dos neurônios no córtex é importante durante a geração dos potenciais elétricos que são registrados pelos eletrodos, pois populações de sinapses são ativadas de diferentes formas em camadas distintas, gerando diferentes padrões de potenciais (SCHOME; SILVA, 2010). Sendo assim, todos os processos neurológicos são dependentes de uma interação complexa entre células e neurônios.

### 2.1.1 Propriedades das células nervosas

O cérebro humano é uma rede com mais de 100 bilhões de neurônios interconectados em sistemas que constroem a nossa percepção do mundo externo (KANDEL et al., 2012). Os neurônios são elementos essenciais para o sistema nervoso e são um tipo de célula altamente especializada.

Um neurônio típico possui quatro regiões morfológicamente definidas, e cada uma destas regiões têm um papel distinto na geração de sinais e na comunicação com outras células nervosas. Estas regiões são: (1) Corpo Celular, ou Soma, (2) Dendritos, (3) Axônios e (4) Terminais Pré-sinápticos, ou terminal do axônio (OLIVEIRA, 2006). A Figura 5 mostra uma estrutura de um neurônio biológico típico.

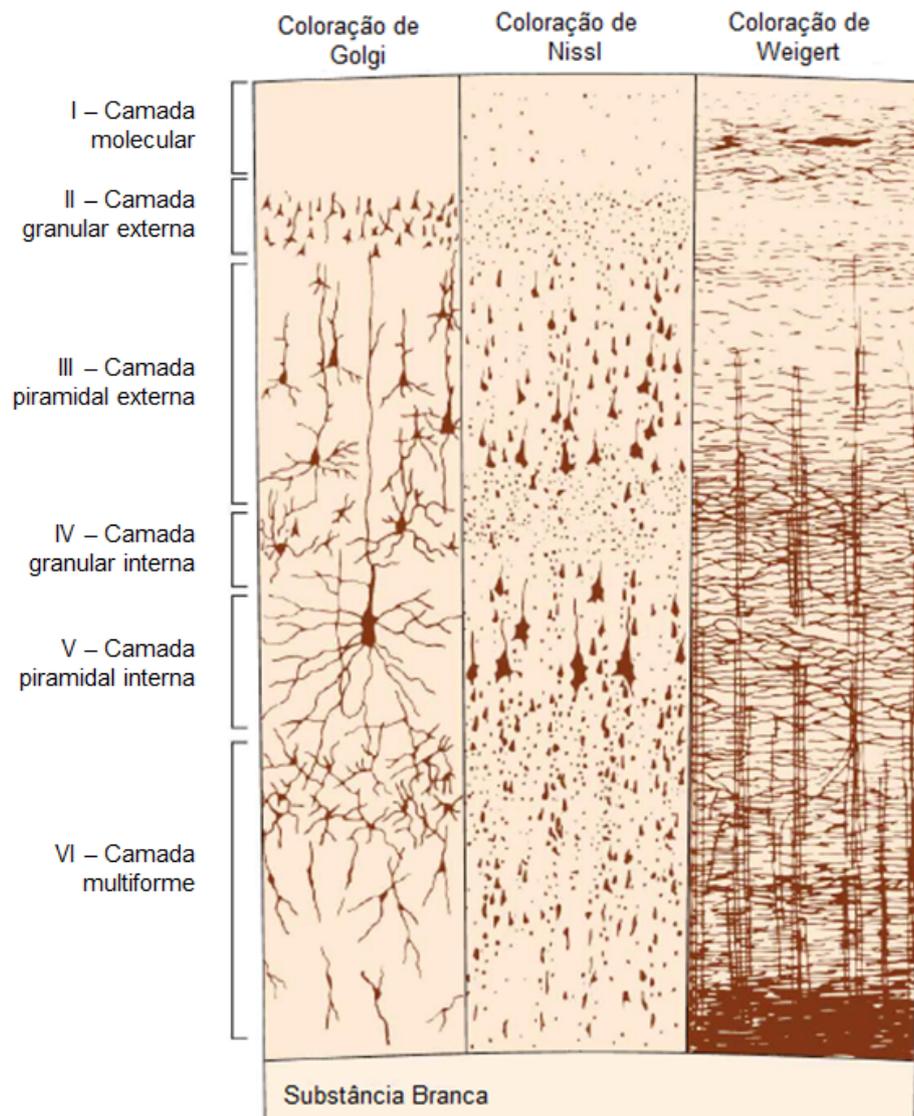
O Corpo Celular, ou Soma, é o centro metabólico do neurônio, onde estão concentradas as principais organelas intracelulares. Os dendritos recebem sinais vindos de outras células nervosas, enquanto o axônio transmite sinais a outros neurônios (KANDEL et al., 2012). Os neurônios podem ser classificados de diferentes formas: de acordo com a função (sensores, motores ou interneurônios), localização (corticais, espinhais, etc.), tipo de transmissor que sintetizam e liberam (glutamatérgicos, colinérgicos, etc.) e forma geométrica (piramidais, granulares, mitrais, etc.) (SQUIRE et al., 2008).

O conteúdo interno do neurônio é separado do meio externo por uma membrana com bicamada fosfolipídica cuja permeabilidade seletiva permite a troca de íons. Esta é a propriedade que torna possível a geração e propagação de sinais bioelétricos (KREBS; WEIBERG; AKES-SON, 2011). Sendo assim, os canais iônicos que se estendem pela membrana são formados

<sup>1</sup>Grande parte do neocórtex apresenta 6 camadas que são numeradas da pia-máter (superfície mais externa do córtex) até a substância branca. A aparência do neocórtex depende da técnica utilizada, como apresentado na Figura 4 (KANDEL et al., 2012).

<sup>2</sup>Para estudar a estrutura das células nervosas foram desenvolvidas diferentes técnicas de coloração que permitem corar seletivamente as células do tecido nervoso (BEAR; CONNORS; PARADISO, 2015).

Figura 4 – Neurônios do neocórtex arranjados em diferentes camadas utilizando diferentes técnicas de coloração.

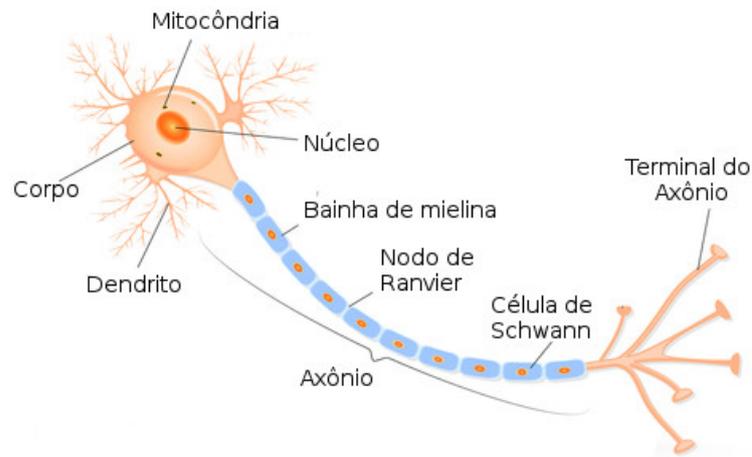


Fonte: Adaptado de Kandel et al., 2012.

por moléculas proteicas que fornecem caminhos para os íons cruzarem a membrana neuronal. Nesses canais iônicos existem “portões” que podem ser abertos ou fechados por meio de alterações locais na membrana, permitindo que haja um caminho entre os meios intra e extracelular. Exemplificando, canais de potássio são seletivamente permeáveis ao íon de potássio ( $K^+$ ), canais de sódio ao íon de sódio ( $Na^+$ ) e assim por diante (BEAR; CONNORS; PARADISO, 2015).

A região intracelular possui uma carga elétrica negativa comparada com a extracelular. Essa diferença de cargas na membrana é chamada de Potencial de Repouso que, em um neurônio típico, é cerca de -65 mV (BEAR; CONNORS; PARADISO, 2015). O movimento de íons

Figura 5 – Estrutura de um neurônio biológico.



Fonte: Oliveira, 2006.

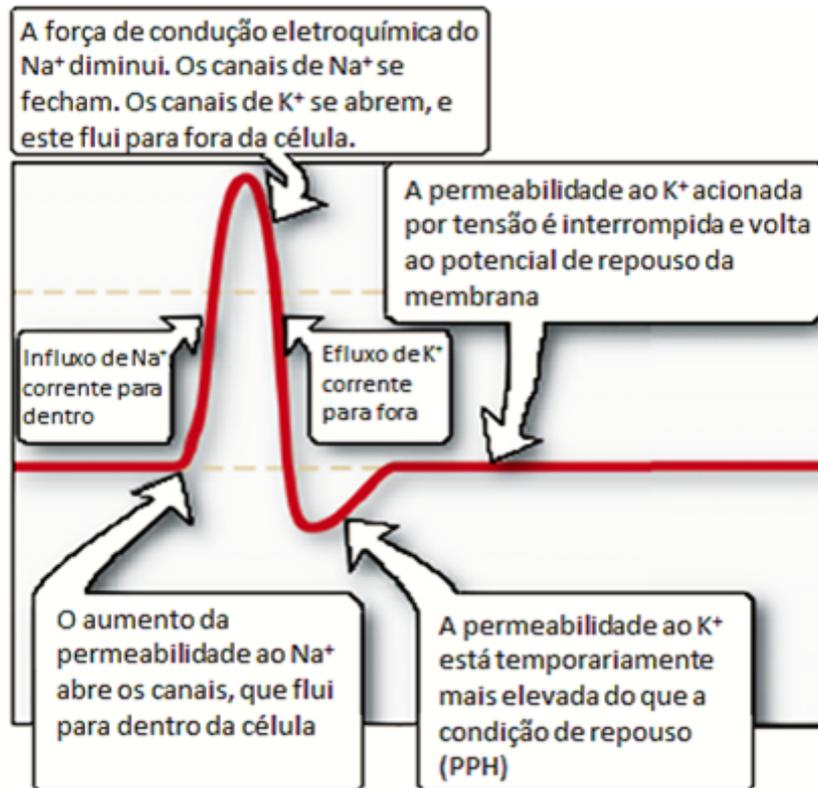
através da membrana gera um gradiente elétrico para cada íon. A soma de todos esses gradientes é o Potencial de Membrana ( $V_m$ ) (KREBS; WEIBERG; AKESSON, 2011). Essa capacidade da membrana de alterar ativamente a sua permeabilidade para diferentes íons (por meio dos canais iônicos) gera um fluxo de corrente que altera o potencial da membrana, fazendo com que seu interior se torne positivamente carregado em relação ao exterior, gerando um Potencial de Ação (PA) (BEAR; CONNORS; PARADISO, 2015). A movimentação dos íons pode causar a hiperpolarização ou despolarização da membrana.

Os PAs são, portanto, impulsos elétricos (ou alterações no potencial da membrana), dependentes de proteínas especiais que atravessam a bicamada fosfolipídica, sendo causados pela despolarização da membrana além do limiar (KREBS; WEIBERG; AKESSON, 2011). São inicializados em uma região próxima da origem do axônio, o Cone de Inserção, cuja amplitude do sinal elétrico é constante ( $\approx 100$  mV), sendo uma resposta do tipo “tudo ou nada” (KANDEL et al., 2012). Os PAs ocorrem apenas nos axônios e são sempre iguais no mesmo neurônio, mas, devido à variabilidade entre as células, são diferentes de um neurônio para outro, mudando a sua velocidade de transmissão, quantidade de canais, etc.

Sabe-se que o íon de potássio ( $K^+$ ) está mais concentrado no meio intracelular, enquanto o íon de sódio ( $Na^+$ ) e cálcio ( $Ca^{2+}$ ) estão mais concentrados no meio extracelular. A entrada de sódio no meio intracelular despolariza a membrana (superfície interna da membrana se torna menos negativa) em resposta a mudanças no seu potencial. Quando essa despolarização atinge um determinado limiar, ocorrerá um PA (KANDEL et al., 2012), isto é, ocorre uma redistribuição de cargas elétricas na membrana, conforme ilustrado na Figura 6.

Após atingir o limiar ocorre a repolarização, provocada pela saída de íons de potássio do meio intracelular, até o interior se tornar negativamente carregado (levando o potencial de membrana de 0 mV a -80 mV), hiperpolarizando a membrana. Como o potencial de membrana se torna mais negativo em relação ao potencial de equilíbrio, os íons de  $Na^+$  tornam a entrar no meio intracelular, atingindo então o potencial de equilíbrio de aproximadamente -65 mV.

Figura 6 – O potencial de ação.



Fonte: Adaptado de Krebs, Weiberg e Akesson, 2011, p.11.

O potencial de equilíbrio é determinado pela (1) concentração de íons dentro e fora da célula, (2) temperatura da solução, (3) a valência do íon e (4) a quantidade de trabalho requerida para separar uma determinada quantidade de carga (SQUIRE et al., 2008), sendo calculado pela Equação de Nernst:

$$E_{ion} = 2.303 \left( \frac{R \times T}{z \times F} \right) \times \log \left( \frac{[ion]_o}{[ion]_i} \right), \quad (1)$$

em que  $E_{ion}$  é o potencial de equilíbrio,  $R$  é a constante dos gases,  $T$  é a temperatura absoluta,  $z$  é a carga do íon,  $F$  é a constante de Faraday,  $\log$  é o logaritmo na base 10,  $[ion]_o$  é a concentração iônica extracelular e  $[ion]_i$  é a concentração iônica intracelular (BEAR; CONNORS; PARADISO, 2015).

Após a ocorrência do PA no neurônio pré-sináptico a informação transmitida ao longo do axônio é enviada ao neurônio pós-sináptico por meio da Transmissão Sináptica, formando uma rede funcional para o processamento e armazenamento de informações (KREBS; WEIBERG; AKESSON, 2011). Portanto, os neurônios se comunicam entre si por meio das sinapses e a transmissão sináptica é fundamental para as funções neurais como a percepção, aprendizado, movimento voluntário, etc. (KANDEL et al., 2012).

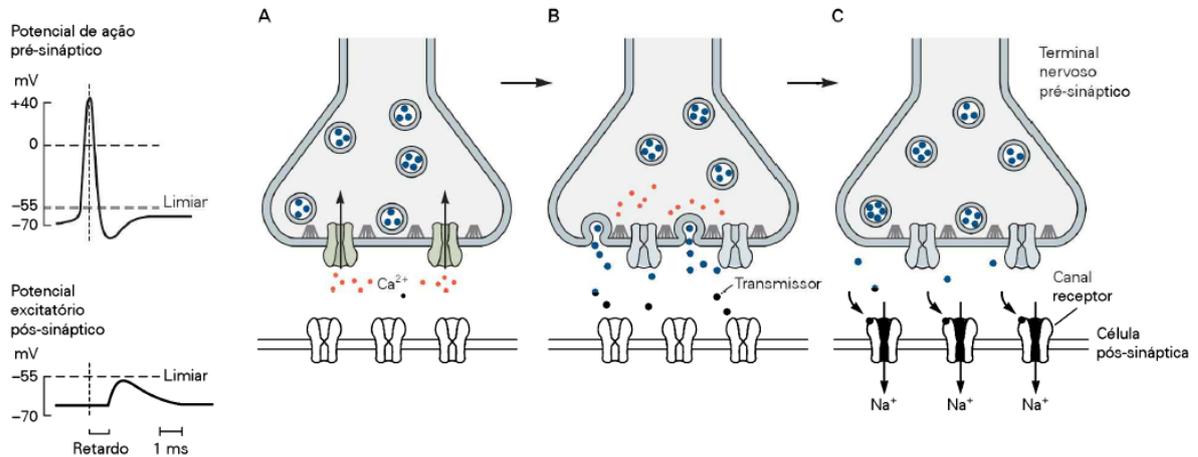
A sinapse apresenta dois lados, o pré e o pós-sinápticos, que indicam a direção do fluxo de informação. As sinapses podem ser do tipo elétrica ou química, sendo que as elétricas possuem uma resposta rápida e de curta duração, tendo uma comunicação citoplasmática com uma distância entre o neurônio pré e pós-sináptico de apenas 4 nm, enquanto as químicas apresentam uma resposta mais lenta, dependendo da difusão do neurotransmissor na fenda sináptica (espaço entre as células pré e pós-sinápticas), com uma distância entre o neurônio pré e pós-sináptico de 20 a 40 nm (KANDEL et al., 2012). Como a maioria das sinapses no cérebro humano é do tipo química, tais mecanismos são descritos mais detalhadamente a seguir.

A Figura 7 descreve o processo da transmissão sináptica química, sendo possível verificar que (A) o PA no terminal pré-sináptico causa a abertura de canais de cálcio dependentes de voltagem na zona ativa, onde as vesículas contendo neurotransmissores estão agrupadas, permitindo a entrada dos íons de cálcio no terminal. O aumento da concentração de íons de cálcio ( $Ca^{2+}$ ) (B) no terminal pré-sináptico faz com que as vesículas sinápticas se unam à membrana celular, liberando neurotransmissores na fenda sináptica por exocitose. O neurotransmissor atravessa a fenda sináptica (C), causando a abertura dos canais iônicos no terminal pós-sináptico, alterando a condutância da membrana e do potencial da célula pós-sináptica (KANDEL et al., 2012).

Quando a liberação de neurotransmissores causa uma despolarização na membrana pós-sináptica, ocorre então um Potencial Excitatório Pós-Sináptico (PEPS). A abertura de canais iônicos por acetilcolina e glutamato causam PEPS. Quando a liberação de neurotransmissores causa uma hiperpolarização na membrana pós-sináptica, ocorre então um Potencial Inibitório Pós-Sináptico (PIPS). A abertura de canais iônicos por glicina ou GABA causam PIPS (BEAR; CONNORS; PARADISO, 2015).

Uma propriedade importante das sinapses químicas é a amplificação, na qual uma vesícula sináptica é capaz de liberar milhares de neurotransmissores que podem abrir vários canais iônicos na célula-alvo. O aumento de neurotransmissores na fenda sináptica resulta em um

Figura 7 – Transmissão sináptica química.



Fonte: Adaptado de Kandel et al., 2012, p.185.

aumento na ativação do receptor, consequentemente gerando maior fluxo de íons ao terminal pós-sináptico (KREBS; WEIBERG; AKESSON, 2011).

A sinalização entre os neurônios ocorre quando os neurotransmissores se ligam aos receptores pós-sinápticos (KANDEL et al., 2012). É o receptor que determina a ação do neurotransmissor, sendo que o neurotransmissor pode afetar direta (receptores ionotrópicos) ou indiretamente (receptores metabotrópicos) os canais iônicos (KREBS; WEIBERG; AKESSON, 2011). Estes receptores são diferentes em seus mecanismos bioquímicos, duração da ação e funções fisiológicas.

Receptores ionotrópicos possuem uma ação sináptica mais rápida, mediando a maioria dos processos motores e sensoriais. Em contrapartida, os receptores metabotrópicos possuem ações mais lentas, modulando a força e eficiência da transmissão sináptica, mediando o comportamento e fortalecendo os caminhos no processo de aprendizagem (KANDEL et al., 2012).

Essas propriedades individuais das células nervosas são fundamentais para a compreensão do funcionamento de grandes redes neurais que permitem a geração das funções cognitivas. Com a descoberta e compreensão da eletricidade e de técnicas experimentais, foi possível realizar o registro e análise destas atividades neurais, permitindo conhecer a natureza da ação dos músculos e nervos (SQUIRE et al., 2008).

## 2.2 GERAÇÃO E DINÂMICA DOS SINAIS DE EEG

O Eletroencefalograma (EEG) é uma técnica não-invasiva que realiza o registro do potencial elétrico do cérebro por meio de eletrodos fixados no escalpo. Seu surgimento ocorreu em 1875, por meio do trabalho realizado por Richard Caton, registrando as correntes elétricas na superfície exposta do encéfalo de cães e coelhos. Posteriormente, em 1929, Hans Berger realizou o primeiro registro de EEG em humanos (BEAR; CONNORS; PARADISO, 2015).

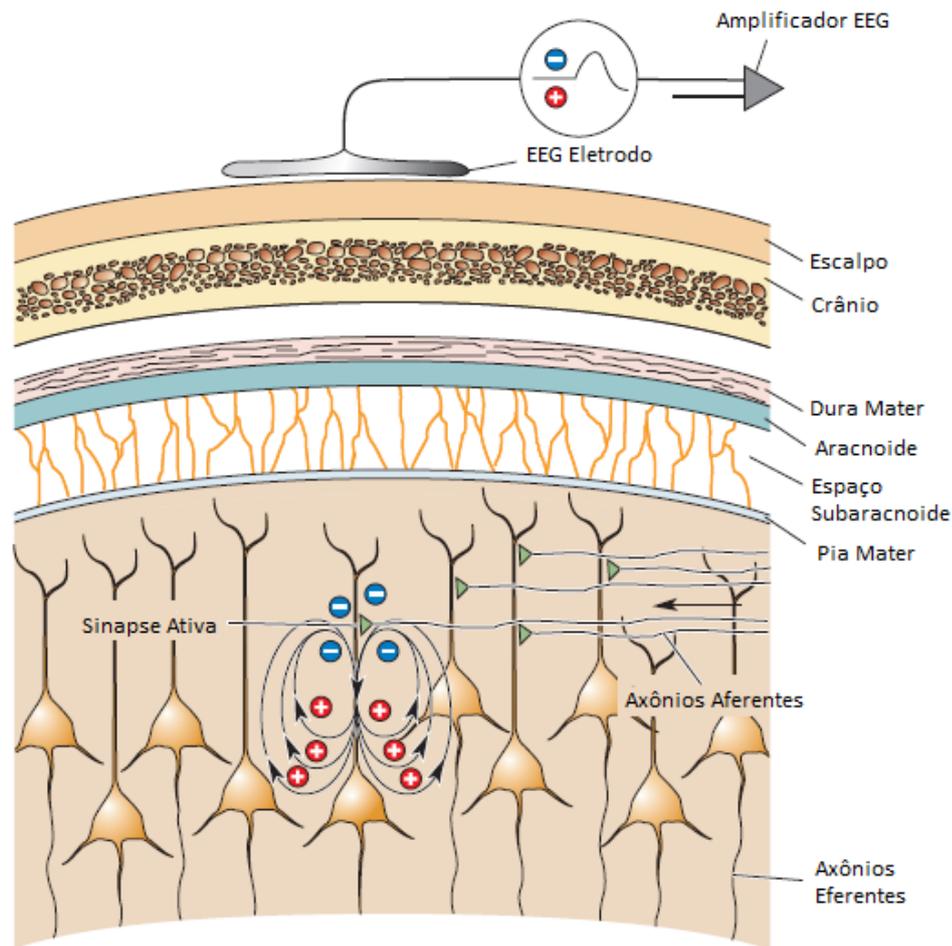
Sabe-se que as funções do córtex são determinadas pelo comportamento coletivo de grupos de neurônios, em vez da ação independente de neurônios individuais (KROPOTOV, 2008). São necessários milhares de neurônios ativados em conjunto para gerar um sinal elétrico grande o suficiente para ser registrado pelo EEG, refletindo um comportamento médio desses neurônios (BUZSÁKI, 2006). Qualquer membrana excitável pode contribuir com a medida de campo do potencial elétrico, sendo que as correntes extracelulares de vários elementos individuais devem se sobrepor para produzirem um sinal grande o suficiente para ser registrado pelos eletrodos (BUZSÁKI; ANASTASSIOU; KOCH, 2012). A amplitude do sinal registrado depende, portanto, de quão síncrona é a atividade dos neurônios subjacentes (BEAR; CONNORS; PARADISO, 2015).

As correntes elétricas registradas pelo EEG de escalpo amostram principalmente as atividades sinápticas que ocorrem nas camadas superficiais do córtex, quando axônios aferentes ipsilaterais (mesmo lado) fazem sinapse com dendritos dos neurônios localizados nessa região (BEAR; CONNORS; PARADISO, 2015), como mostra a Figura 8.

Eventos lentos, da ordem de 50 a 200 ms, como as correntes sinápticas, são mais propícios a serem medidos pelo EEG. A somação dos PIPS e PEPS na rede neural forma correntes elétricas que fluem dentro e ao redor das células, criando um campo que se espalha ao redor da origem do sinal elétrico (MARCUSE; FIELDS; YOO, 2016), formando um dipolo. O efeito deste campo diminui conforme se aumenta a distância entre a fonte do sinal e o eletrodo.

Como existem diferentes tipos de neurônios, as pequenas correntes geradas nos dendritos de um mesmo neurônio são somadas de maneiras distintas. Neurônios do tipo piramidal formam dipolos propícios para o registro dos sinais pelos eletrodos. Dessa forma, é possível dizer que o EEG registra a soma de milhares de pequenos dipolos resultantes do comportamento elétrico de neurônios piramidais orientados de maneira perpendicular à superfície do escalpo (KROPOTOV, 2008).

Figura 8 – Geração do Sinal Elétrico do EEG.



Fonte: Adaptado de Bear, Connors e Paradiso, 2006, p.587.

O posicionamento dos eletrodos deve seguir a convenção 10-20 recomendada pela *International Federation of Societies for Electroencephalography and Clinical Neurophysiology*, sendo posicionados de acordo com marcos anatômicos determinados, sendo 10% ou 20% o intervalo dos eletrodos (BEAR; CONNORS; PARADISO, 2015).

Existem duas formas para registrar os sinais de EEG, sendo essas o Registro Bipolar, em que é computada a diferença de potencial entre dois eletrodos posicionados no escalpo, e o Registro Referencial, em que é computada a diferença de potencial entre um eletrodo e um ponto de referência que não é afetado pelos potenciais gerados pela atividade cerebral como, por exemplo, as orelhas.

As oscilações são componentes essenciais do EEG, geradas por grupos de neurônios, caracterizadas pela sua frequência, amplitude e fase. Os ritmos cerebrais e a sincronização estão relacionados com funções cognitivas como a transferência de informação, percepção, controle

motor e memória (KAMEL; MALIK, 2014). O sinal de EEG pode ser caracterizado em termos de bandas de frequência que podem ser brevemente descritas como (SCHOME; SILVA, 2010):

- a) Delta (0,5 a 4 Hz): Os ritmos delta estão relacionados com o sono e a anestesia, sendo que a frequência dessas oscilações lentas depende da profundidade ou do tipo de anestesia, ou do estágio do sono (SCHOME; SILVA, 2010). Podem ter origem cortical ou talâmica (KROPOTOV, 2008). As oscilações delta sinalizam uma maior redução nos processos de excitação encefálica e um aumento na sincronização da atividade cortical e talâmica (KANDEL et al., 2012);
- b) Teta (4 a 8 Hz): As oscilações teta ocorrem durante a solução de problemas (KROPOTOV, 2008). Assim, os ritmos teta podem representar estados dinâmicos que emergem de redes hipocâmpais engajadas na navegação espacial e em processos da memória (SCHOME; SILVA, 2010);
- c) Alfa (8 a 12 Hz): Os ritmos alfa occipitais (córtex visual) ocorrem durante a atenção visual reduzida e os ritmos *mu* (córtex somatossensorial) ocorrem em um estado de relaxamento muscular (SCHOME; SILVA, 2010), estando relacionados com um estado ocioso das áreas corticais correspondentes (KROPOTOV, 2008);
- d) Beta (12 a 30 Hz): Os ritmos beta ocorrem em estado de consciência/vigília, estando associados com a concentração durante solução de problemas e tomadas de decisão (KROPOTOV, 2008);
- e) Gama (>30 Hz): As oscilações gama refletem um estágio de processamento ativo de informações (SCHOME; SILVA, 2010) relacionadas a um alto nível de processamento cognitivo.

A vantagem do EEG, frente a outros equipamentos, é a sua resolução temporal, da ordem de milissegundos. Os potenciais elétricos evocados a partir de um determinado estímulo é uma técnica comumente empregada para se analisar esses sinais cerebrais, como descrito a seguir.

### 2.2.1 Potencial relacionado ao evento

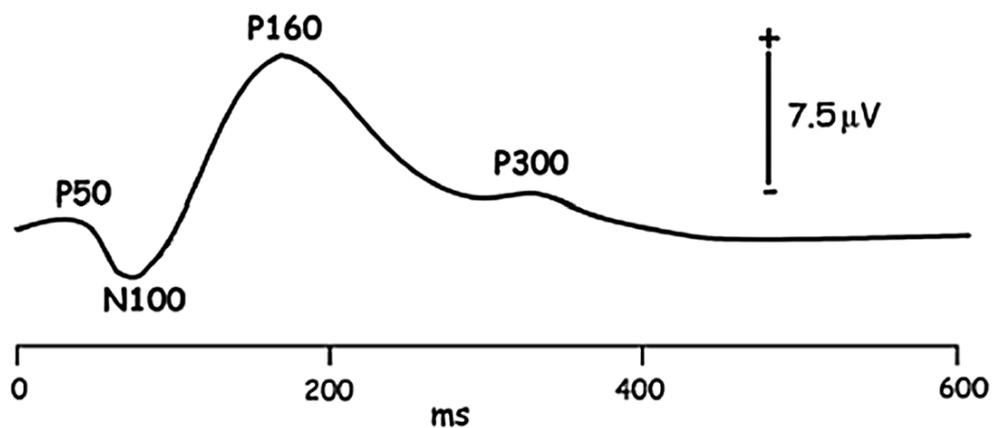
A cada momento o cérebro apresenta um estado funcional resultante das novas informações recebidas. Conseqüentemente, inicia-se uma série de diferentes e breves estados cerebrais que manifestam distintos aspectos do processamento de tais informações (LEHMANN, 1989).

O Potencial Relacionado ao Evento (do inglês, *Event-Related Potential* - ERP) é uma técnica que registra os potenciais elétricos que são evocados a partir de um determinado estí-

mulo. Ou seja, quando são apresentados determinados estímulos, é possível observar no sinal de EEG a geração de curvas com picos proeminentes em resposta, refletindo o processamento cerebral. Este comportamento reflete quase sempre a resposta da transmissão sináptica em vez dos PAs. No entanto, as respostas auditivas evocadas no tronco encefálico são uma exceção à regra, pois refletem o disparo sincronizado de PAs que são da ordem de poucos milissegundos (LUCK, 2014). O ERP permite a observação de uma série de operações cognitivas que ocorrem quando uma informação sensorial é enviada pelo sistema nervoso central, antes mesmo que uma resposta comportamental seja realizada (WOODMAN, 2010).

Aos diferentes aspectos do processamento das novas informações apresentadas ao cérebro é dado o nome de “componentes” (LEHMANN, 1989), caracterizados pela sua latência, morfologia e topografia (KAMEL; MALIK, 2014). Convencionalmente, os componentes de ERP são representados pela letra N quando sua polaridade é negativa e P quando sua polaridade é positiva, seguidos por um número que representa a posição temporal do pico na forma de onda (e.g. N300 para um pico negativo em 300 ms) ou a sua posição ordinal (e.g. N1 para o primeiro pico negativo), conforme demonstrado na Figura 9 (LUCK, 2014).

Figura 9 – Sequência típica dos componentes ERP a partir da apresentação de estímulos auditivos.



Fonte: Adaptado de Luck, 2014, p.80.

Um componente de ERP é normalmente avaliado pelo potencial elétrico registrado em seu pico ou pela área abaixo da curva em uma pequena janela de tempo ao redor do seu pico. Como exemplo, a latência do pico pode ser interpretada como a relação entre o tempo do processamento neural para um determinado estímulo, considerada um método simples, visual e intuitivo (BOSTANOV, 2015).

Os componentes de ERP podem ser divididos em 3 categorias principais: (1) Componentes exógenos, ativados pela presença de um estímulo; (2) Componentes endógenos, que refletem o processamento neural dependente de uma determinada tarefa; e (3) Componentes motores: relacionados com a preparação e execução de uma resposta motora (LUCK, 2014).

Os potenciais evocados auditivos estão relacionados com uma atividade neural específica, resultante de uma estimulação acústica. Esses potenciais podem ser caracterizados como: (1) Potenciais de curta latência, que surgem normalmente nos primeiros 10 ms pós-estímulo, refletindo o fluxo de informação vindo da cóclea ao tronco encefálico e tálamo, normalmente utilizados para analisar a integridade da via auditiva; (2) Potenciais de média latência, posteriores aos de curta latência até aproximadamente 80 ms pós-estímulo, refletindo o fluxo de informação entre o tálamo e o córtex, utilizado para identificar a sensibilidade auditiva; e (3) Potenciais de longa latência ( $> 80$  ms), que possuem origem cortical e são altamente dependentes do tipo de estímulo utilizado, local de registro, técnica de registro, idade e estado do sujeito (KRAUS; NICOL, 2008).

Alterações na morfologia dos potenciais evocados auditivos corticais, como o aumento da amplitude e diminuição da latência, indicam aumento no sincronismo neural e fortalecimento das conexões neurais (SANJU; KUMAR, 2016). Desta forma, os potenciais de longa latência podem ser influenciados por fatores de alto nível, como a atenção e a excitação (e.g. o componente N1 que reflete um estado de atenção ao estímulo apresentado).

Os potenciais de longa latência considerados como componentes exógenos se iniciam pelo P1 (muitas vezes classificado como um potencial de média latência), ocorrendo em aproximadamente 80 ms pós-estímulo, até o componente N2 (aproximadamente 250 ms pós-estímulo). Alterações na frequência, complexidade e intensidade dos tons são estímulos típicos para a ocorrência desses componentes (KRAUS; NICOL, 2008).

Entre os componentes exógenos e endógenos está o *mismatch negativity* (MMN), considerado como um detector de mudanças acústicas (LUCK, 2014). O MMN é evocado por uma sequência de sons idênticos interrompida ocasionalmente por um som discrepante (denominado paradigma *oddball*) (KRAUS; NICOL, 2008). O MMN é comumente utilizado para acessar as habilidades de discriminação auditivas pré-atencionais que ocorrem de maneira inconsciente (SANJU; KUMAR, 2016).

Dos componentes endógenos, especificamente o P3, evocado por meio do paradigma *oddball*, ocorre quando o ouvinte conscientemente atende ao estímulo discrepante realizando alguma tarefa específica, sendo considerado como um indicador de cognição por serem neces-

sários uma avaliação e classificação do estímulo apresentado. Já o componente N4 ocorre em resposta a estímulos da fala, quando ocorre alguma incongruência semântica, sendo um indicador do processamento da linguagem, requerendo um alto nível de cognição (KRAUS; NICOL, 2008).

Resumidamente, a técnica de ERP permite encontrar diferenças súbitas nos campos eletromagnéticos que geram mudanças neurais rápidas em múltiplas localizações espaciais com resolução de milissegundos (GROPPE; URBACH; KUTAS, 2011).

### 2.3 CONSIDERAÇÕES COMPLEMENTARES

É de fundamental importância compreender as estruturas que compõem o sistema nervoso e como os seus mecanismos são capazes de fazer o ser humano interagir com o meio em que se encontra. Os neurônios formam uma rede neural complexa responsável por diversas funções que garantem o funcionamento do nosso organismo. Contudo, pequenas falhas nessa rede neural são capazes de causar danos significativos na integridade do sistema cognitivo. Desta forma, é necessário encontrar meios para se estudar os diversos padrões de ativação neurais responsáveis pela execução de funções como o pensamento, a memória, o aprendizado, entre outros.

Nesta tese é descrito e utilizado o EEG por auxiliar nessa compreensão. Resultados recentes e interessantes têm sido encontrados com relação ao processamento da música utilizando o EEG, como será descrito a seguir, demonstrando por meio da técnica de ERP os distintos aspectos funcionais do processamento humano de tais informações.

### 3 A MÚSICA E O CÉREBRO

Para tocar um instrumento musical são necessárias diversas ações executadas paralelamente, como o ato de ouvir, planejar, sentir, mover, coordenar e lembrar os diversos elementos musicais. A prática musical é, portanto, uma das experiências humanas mais ricas por envolver tanto a ativação emocional, como a integração sensorial, motora e a atividade cognitiva (ALTENMÜLLER; SCHLAUG, 2015). Uma boa performance musical requer um controle de movimentos complexos sob contínuo *feedback* auditivo. Sendo assim, é necessário uma integração audio-motora para o processamento musical, o que torna essa prática uma intensa experiência multisensorial (ZATORRE; CHEN; PENHUNE, 2007).

O cérebro humano possui uma grande capacidade de se auto-organizar em resposta a estímulos externos (RAUSCHECKER, 2003). A experiência musical ao longo da vida produz impactos significativos nas funções cognitivas, oferecendo uma oportunidade única para compreender como o cérebro humano se altera em resposta ao treinamento e a aquisição de novas habilidades (HUTCHINSON et al., 2003; CHEN; PENHUNE; ZATORRE, 2008), além de possuir um papel importante na exploração da natureza e da extensão da plasticidade cerebral (PERETZ; ZATORRE, 2005), afinal, os músicos estão constantemente envolvidos com este estímulo complexo e são expostos a própria música por longo tempo (MÜNTE; ALTENMÜLLER; JÄNCKE, 2002; SCHLAUG, 2015).

A prática musical pode apresentar diversos benefícios, tais como a capacidade de promover um aumento na interconexão entre os dois hemisférios cerebrais (SCHLAUG et al., 1995), proteção contra o declínio cognitivo relacionado ao envelhecimento (EGGERMONT, 2014; KOK, 2000; BAIRD; SAMSON, 2015; SARKAMO, 2017; O'BRIEN; NIKJEH; LISTER, 2015), benefícios para habilidades linguísticas (SARKAMO, 2017; CARPENTIER; MORENO; MCINTOSH, 2016; ALTENMÜLLER; SCHLAUG, 2015; BIDELMAN; ALAIN, 2015), assim como para fins terapêuticos, regulando o humor, evocando memórias, realizando a manutenção da autoestima, reduzindo sentimentos de solidão e isolamento, entre outros (SARKAMO, 2017; O'BRIEN; NIKJEH; LISTER, 2015; DUSHANOVA; CHRISTOV, 2014; HO et al., 2012; ALTENMÜLLER; SCHLAUG, 2015). Esses benefícios da música conferem ao músico uma melhor habilidade para a organização perceptual dos sons (ALAIN et al., 2014). Em outras palavras, a música traz uma melhor qualidade de vida para todos e, em particular, para quem a pratica.

O fato de a atividade musical estimular redes neurais distribuídas que estão intimamente ligadas com muitos sistemas neuroquímicos, regulando a emoção, excitação e funções cognitivas, faz da música uma potente ferramenta para ser utilizada na reabilitação neurológica (SARKAMO, 2017). Evidências apontam que o estilo de vida impacta diretamente no envelhecimento (BARTOLOTTI; LAZAROV, 2016). Assim, participar de atividades físicas, estar socialmente ativo e participar de atividades cognitivamente estimulantes (e.g. leitura, aprendizado, jogos) podem reduzir o risco de degeneração cognitiva ou de demência (ALAIN et al., 2014). Existem, portanto, expectativas quanto ao potencial das intervenções baseadas em música para promover reservas cognitivas, além de proporcionar bem estar emocional durante o envelhecimento normal (BAIRD; SAMSON, 2015; SCHLAUG, 2015).

Neste capítulo serão descritos os mecanismos para o processamento dos sons, em particular da música, destacando as diferenças existentes entre o cérebro de músicos e não-músicos. Ao final será feita uma breve análise crítica dos benefícios da música para o envelhecimento, além de uma descrição sucinta sobre reserva cognitiva.

### 3.1 COMO O CÉREBRO PROCESSA A MÚSICA

As vibrações sonoras são capazes de estimular o sistema auditivo provocando uma percepção sonora. Uma série de eventos deve acontecer para que o cérebro processe a música. Isso ocorre por meio de um longo circuito auditivo, a fim de formar uma percepção auditiva, originando-se nos órgãos sensoriais, se dirigindo para o tronco encefálico e o tálamo, até atingir o córtex auditivo primário, distribuindo a informação para as áreas associativas do córtex (BEAR; CONNORS; PARADISO, 2015).

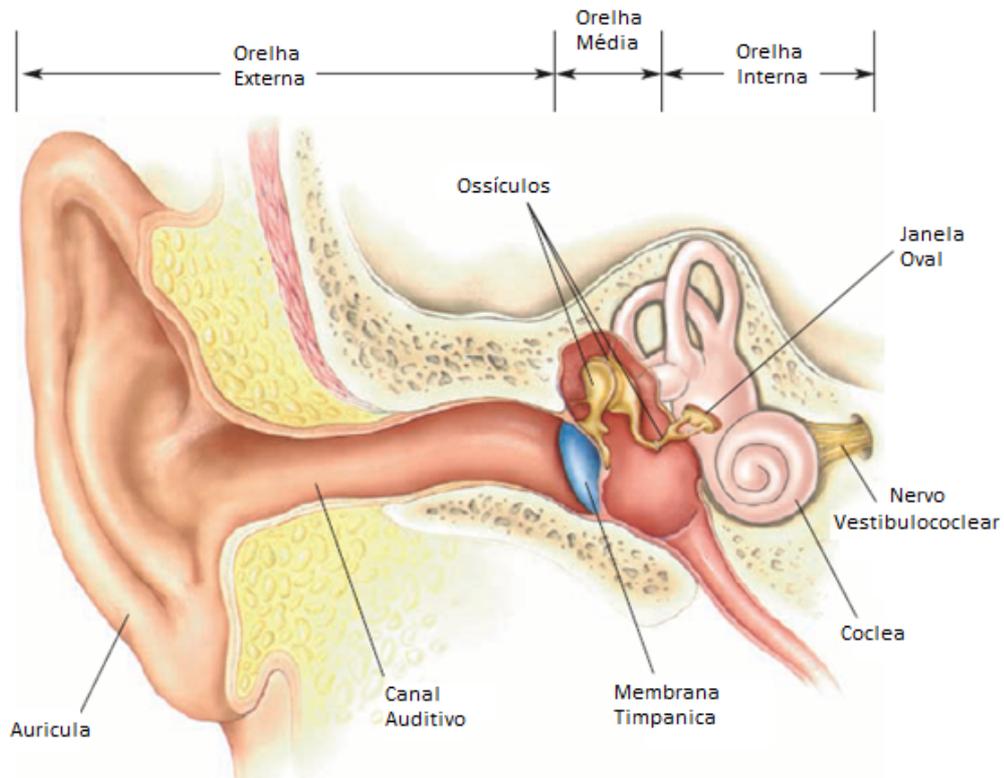
Quando um estímulo sonoro incide sobre o tímpano, este começa a vibrar em resposta, transmitindo essa vibração para os três ossículos articulados existentes dentro da orelha média: o martelo, a bigorna e o estribo (BEAR; CONNORS; PARADISO, 2015). Quando o estribo se move, a membrana basilar (existente dentro da cóclea) se move em resposta, ativando os mecanorreceptores (células ciliadas) responsáveis por transformar a energia mecânica resultante em um sinal interpretável pelo cérebro (KANDEL et al., 2012). Há, inclusive, uma relação tonotópica<sup>1</sup> entre a frequência de vibração do som e a localização das células ciliadas na membrana

---

<sup>1</sup>Tonotopia se refere a relação entre a frequência sonora e a posição das células ciliadas ao longo da membrana basilar na cóclea (KANDEL et al., 2012). Ou seja, na membrana basilar é formado um mapa arranjado de tal forma que uma determinada área corresponde a uma frequência em particular.

basilar (BEAR; CONNORS; PARADISO, 2015). Dessa forma, a cóclea transforma a energia sonora em sinais elétricos e os envia ao encéfalo, conforme ilustrado na Figura 10.

Figura 10 – Ilustração do circuito auditivo.



Fonte: Adaptado de Bear, Connors e Paradiso, 2006, p.347.

Neurônios estabelecem sinapses com as células ciliadas, projetando seus axônios no nervo vestibulo-coclear (VIII nervo craniano) que leva a informação sonora de maneira organizada ao córtex auditivo (A1 ou área 41 de Broadmann), localizado no Lobo Temporal, cujas aferências fazem sinapses com as camadas corticais IIIb e IV (BEAR; CONNORS; PARADISO, 2015). Sendo assim, uma vez que o som foi transformado em uma resposta elétrica na cóclea, a percepção auditiva é gerada a partir de uma rica hierarquia de circuitos auditivos que analisam e processam estes sinais (KANDEL et al., 2012).

O córtex auditivo se conecta com o restante do encéfalo por meio de vários circuitos em um sistema hierárquico (ZATORRE; CHEN; PENHUNE, 2007) composto por estruturas como: o hipocampo, que é uma área relacionada à memória e permite reconhecer a familiaridade das melodias e ritmos; o cerebelo e a amígdala, relacionados respectivamente com a regulação motora e emocional; o núcleo acúmbens, um pequeno núcleo de substância cinzenta relacionado ao sentido do prazer e recompensa (MUSZKAT; CORREIA; CAMPOS, 2000); e áreas do cór-

tex motor e parietal que estão relacionadas com a percepção do ritmo e com a ordem temporal (PERETZ; ZATORRE, 2005). Ainda assim, a via auditiva é longa e há muito para ser compreendido sobre o seu funcionamento (KANDEL et al., 2012). Seu circuito auditivo é capaz de transformar uma onda sonora, puramente mecânica, em uma percepção auditiva, respondendo a um código de localização sonoro (tonotopia), permitindo ao ser humano interagir com o meio em que se encontra.

Quando se fala sobre o processamento dos sons, é importante conhecer a definição de discriminação de tons (do inglês, *pitch discrimination*) e o tom<sup>2</sup>, ou altura (do inglês, *pitch*), propriamente dito. O primeiro se refere a habilidade de detectar a menor variação na frequência fundamental ( $f_0$ ) de dois tons sucessivos. O segundo está relacionado com a percepção da frequência fundamental ( $f_0$ ) dos sons sendo que, por exemplo, a nota dó (C4) central do piano corresponde a frequência 532,25 Hz e a nota lá (A3) corresponde a frequência 440 Hz.

Existem dois meios para se analisar e compreender como ocorre o processamento musical: (1) por meio da análise de neuroimagens realizadas por equipamentos como a Imagem por Ressonância Magnética funcional (do inglês, *functional Magnetic Resonance Imaging* - fMRI), Tomografia por Emissão de Pósitrons (do inglês, *Positron Emission Tomography* - PET), Magnetoencefalografia (do inglês, *Magnetoencephalography* - MEG), Eletroencefalografia (do inglês, *Electroencephalography* - EEG), Espectroscopia funcional em Infravermelho Próximo (do inglês, *functional Near-Infrared Spectroscopy* - fNIRS), entre outros, que permitem a determinação de estruturas comumente envolvidas no processamento musical; (2) por meio do estudo de pacientes com lesões encefálicas, o que permite inferir quais as estruturas necessárias para o processamento da música em decorrência da lesão (GRIFFITHS, 2001).

Estudos demonstram que a organização tonotópica observada na cóclea é preservada em todos os núcleos da via auditiva em direção ao córtex auditivo primário (A1) (PATTERSON et al., 2002). Além disso, a detecção de sons com diferentes frequências (tons puros) depende da integridade da via auditiva ascendente dirigida a pelo menos um dos córtex auditivos (GRIFFITHS, 2001). Dessa forma, a via auditiva e os córtex auditivos primários fornecem mecanismos suficientes para o processamento de características espectro-temporais de sons individuais, sendo o primeiro estágio para o processamento neural da música. O processamento subsequente de padrões mais complexos requer a associação de outras áreas do córtex.

---

<sup>2</sup>Tom pode ter vários significados, desde o intervalo correspondente a diferenças de altura entre duas notas, a referência a escalas maiores e menores, ou a altura correspondente de um determinado som que possui uma frequência determinada. Nesta tese, tom se refere a altura de um som.

O córtex auditivo primário direito está envolvido com a percepção de tons complexos (e.g. identificação de tons cuja fundamental  $f_0$  está ausente) (GRIFFITHS, 2001), enquanto o hemisfério esquerdo está associado com o processamento de informações auditivas rápidas e sequenciais (i.e. a produção e percepção do ritmo) (SAMSON; EHRLÉ; BAULAC, 2001). Sabe-se, também, que o lobo temporal direito está associado com o processamento do timbre (SAMSON, 2003) e tem um papel importante na decodificação da altura das notas (PERETZ; ZATORRE, 2005). Ou seja, existe uma especialização hemisférica, na qual o hemisfério esquerdo está relacionado com a resolução temporal (ritmo e métrica), enquanto o hemisfério direito está relacionado com a resolução espectral (discriminação melódica) (ZATORRE; BELIN, 2001). O processamento musical como um todo se torna relativamente mais assimétrico e com maior atividade no hemisfério direito (ZATORRE; BELIN; PENHUNE, 2002).

Uma importante questão a ser levantada é o fato de que a maioria destes estudos funcionais sobre como o cérebro processa a música são baseados em estímulos construídos artificialmente. Ou seja, esses estudos não representam a música como a mesma é de fato escutada, limitando a abrangência destas pesquisas (ABRAMS et al., 2013), por apresentarem apenas dimensões musicais específicas de forma isolada. Portanto, é necessário analisar o processamento da escuta musical como um processo contínuo, utilizando estímulos naturalísticos, ou seja, músicas completas ou pequenos trechos de músicas, que podem prover informações mais precisas do processamento das características musicais no nosso cérebro (ALLURI et al., 2012). Afinal, a habilidade de integrar componentes da informação musical é um dos aspectos mais importantes da escuta musical (ABRAMS et al., 2013). Assim, os estímulos artificiais falham em capturar as modulações causadas devido ao fluxo contínuo de informações que devem ser processados pelo cérebro para gerar uma percepção auditiva (ALLURI et al., 2013).

Propondo uma nova metodologia para se predizer a evolução temporal da atividade neural em condições de escuta musical naturalística e usando uma combinação de características acústicas com neuroimagens, Alluri et al. (2012) apresentaram a música “*Adios nonino*” de Astor Piazzolla a um grupo de voluntários músicos enquanto eram registrados seus sinais de fMRI. Eles encontraram ativações de áreas do cerebelo para o processamento de características do timbre, além de áreas corticais e subcorticais, motoras e circuitos relacionados a emoção para o processamento de características de tonalidade e do pulso musical. Utilizando a mesma metodologia de Alluri et al. (2012), Alluri et al. (2013) realizaram experimentos apresentando músicas cantadas e instrumentais aos voluntários, encontrando que ambos os tipos de estímulo revelam que ativações nas áreas auditivas, límbicas e motoras podem ser preditas. Além disso,

os resultados apontam para especialização hemisférica para os diferentes tipos de estímulo, sendo que aparecem ativações na amígdala e hipocampo para as músicas cantadas, podendo estar relacionadas a familiaridade ao estímulo.

Em um experimento utilizando também fMRI para se analisar as respostas neurais de 17 não-músicos durante a escuta de 4 sinfonias de William Boyce, Abrams et al. (2013) encontraram que a sincronização entre sujeitos (do inglês, *Inter-Subject Synchronization* - ISS) acontece em diversas regiões do cérebro durante a escuta musical de estímulos naturais. A ISS pôde ser vista em diferentes níveis de hierarquia no processamento da informação, desde estruturas auditivas subcorticais e corticais do lobo temporal, a redes de atenção fronto-parietais e áreas de planejamento motor. Esta sincronização se relaciona com aspectos da estrutura musical (e.g. ritmo, timbre, etc.). Além disso, esse efeito de sincronização é maior para músicas naturais comparada com condições de controle pseudo-musicais. Ou seja, a ISS é fortemente reduzida para estímulos não-musicais.

Resumidamente, a prática musical confere uma integração multisensorial em que diferentes regiões do cérebro agem juntamente para gerar a experiência musical (ALTENMÜLLER; SCHLAUG, 2015; BROWN; ZATORRE; PENHUNE, 2015). Temos, portanto, a relação entre os lobos temporais, occipitais e parietais integrando entradas sensoriais da audição, visão e somatosensorial respectivamente, além do lobo frontal envolvido na atenção, no planejamento e preparo motor, integrando a informação auditiva e motora, além de estar relacionado com habilidades humanas específicas como a imitação e a empatia que estão diretamente relacionadas com a aquisição de habilidades musicais e com a expressividade emocional (ALTENMÜLLER; SCHLAUG, 2015). O cerebelo também possui um papel importante na coordenação motora, estando tipicamente relacionado com o processamento do ritmo (BROWN; ZATORRE; PENHUNE, 2015), e sabe-se que músicos demonstram um recrutamento mais eficiente das regiões motoras do cérebro (CHEN; PENHUNE; ZATORRE, 2008). Além disso, a amígdala e o hipocampo são importantes estruturas para a percepção emocional.

### 3.2 DIFERENÇAS COGNITIVAS ENTRE MÚSICOS E NÃO-MÚSICOS

Ao longo dos anos, diversas pesquisas foram feitas a fim de compreender como o cérebro processa a música (MÜNTE; ALTENMÜLLER; JÄNCKE, 2002; PERETZ; ZATORRE, 2005). Por meio dessas pesquisas foram observadas diversas diferenças neurais entre músicos e não-músicos (MÜNTE; ALTENMÜLLER; JÄNCKE, 2002; TERVANIEMI, 2009). Sabe-se que

a experiência musical é capaz de promover profundas mudanças no cérebro, tanto estruturais quanto funcionais (SCHLAUG, 2001). Estas diferenças entre estes dois grupos são descritas a seguir.

Estruturalmente é possível encontrar diferenças entre músicos e não-músicos, tais como o tamanho do Corpo Caloso, que é uma estrutura responsável pela interação e comunicação inter-hemisférica (SCHLAUG et al., 1995), o aumento no volume da substância cinzenta no giro de Heschl (SCHNEIDER et al., 2002), na área de Broca (SLUMING et al., 2002) e em áreas motoras (M1), somatossensoriais (S1), auditivas (A1), visuais e no córtex pré-motor (GASER; SCHLAUG, 2003a,b), além de um aumento no volume cerebelar, relacionado ao processamento e aprendizado motor (HUTCHINSON et al., 2003; KOENEKE et al., 2004). Evidências apontam que durante o período de desenvolvimento (na infância), pianistas apresentam aumento nas vias de substância branca, sugerindo que o treinamento musical pode induzir a plasticidade nos tratos corticoespinhais e corpo caloso, demonstrando uma relação entre a experiência/treinamento e a plasticidade na substância branca (BENGTSSON et al., 2005), havendo também um aumento na conectividade da substância branca entre o *Planum Temporale* esquerdo e direito, relacionado com o processamento auditivo de sons complexos (ELMER; HANGGI; JANCKE, 2016).

Existem também diferenças entre músicos de acordo com o instrumento que tocam, devido ao fato de que cada instrumento pode requerer demandas cognitivas distintas entre si. Assim, ao se comparar violinistas com pianistas, Bangert e Schlaug (2006) encontraram diferenças no córtex motor entre os dois grupos, correspondente a presença do sinal Omega em ambos os hemisférios para pianistas, mas apenas no hemisfério direito em violinistas no giro pré-central, sugerindo a possibilidade de adaptação no cérebro humano durante o período de aprendizado dependendo do tipo de instrumento tocado. Quando comparados com não-músicos, violinistas destros, por exemplo, apresentam uma assimetria inter-hemisférica significativa, relacionada ao aumento da representação da mão esquerda no córtex somatossensorial primário (S1) e no córtex motor primário (M1), identificados por meio da estimulação magnética transcraniana e do potencial evocado somatossensorial após estimulação elétrica (SCHWENKREIS et al., 2007). Diferenças também podem ser vistas entre trompetistas e pianistas, que estão relacionadas com ativações neurais específicas ao instrumento (GEBEL et al., 2013), havendo um aumento no recrutamento do hemisfério cerebelar posterior-superior, no M1, no S1 e no giro de Heschl esquerdo em trompetistas durante execução do trompete. Ou seja, é possível que estas diferenças estruturais sejam consequência do aumento do uso, seja a mão esquerda para violinistas, a mão

direita para o trompetistas ou ambas as mãos para pianistas, além do uso dos lábios e diafragma (sistema respiratório), no caso dos instrumentos de sopro.

Funcionalmente, o cérebro de músicos e não-músicos respondem de maneira diferente para estímulos auditivos. Estudos apontam que as respostas do ERP são maiores em músicos do que em não-músicos. Por exemplo, é possível observar uma maior resposta do componente N1 a sons de piano para músicos, quando comparados com não-músicos (PANTEV et al., 1998), em particular para estímulos que representam o timbre do instrumento executado pelo músico (PANTEV et al., 2001), o que implica em uma especialização para sons mais familiares aos sujeitos. Há um aumento na amplitude do componente P3 em músicos durante a ocorrência de disparidades em relações intervalares (*oddball*), mas não há diferenças entre músicos e não-músicos para contornos melódicos (TRAINOR; DESJARDINS; ROCKEL, 1999). Existe também um aumento na amplitude nas respostas pré-atencionais como o componente MMN (KOELSCH; SCHRÖGER; TERVANIEMI, 1999) e para os componentes N1c<sup>3</sup> e P2 (SHAHIN et al., 2003), o que demonstra que músicos apresentam melhor discriminação tanto para estímulos que requerem alto nível cognitivo quanto para respostas pré-atencionais. Além disso, como acontece com músicos instrumentistas, cantores apresentam melhor discriminação de tons do que não-músicos (NIKJEH; LISTER; FRISCH, 2008), como revelado pelo componente MMN em que músicos (cantores e instrumentistas) apresentaram o pico do MMN com menor latência do que o grupo de não-músicos. Comparando os grupos, cantores que também são instrumentistas apresentaram latência ainda menor.

Em outras palavras, assim como é possível encontrar diferenças estruturais entre os próprios músicos, funcionalmente também é possível diferenciá-los. Maestros apresentam melhor processamento espacial do que pianistas e não-músicos, além de apresentarem o componente MMN seguido por um pico positivo (P3a<sup>4</sup>), sugerindo uma mudança do foco da atenção em direção ao estímulo desviante, diferente dos outros grupos observados (NAGER et al., 2003). Existem também diferenças no processamento cognitivo de músicos que executam diferentes estilos/gêneros musicais (e.g. clássico, jazz) (VUUST et al., 2012), sendo que músicos de jazz apresentam maior amplitude ao componente MMN do que músicos clássicos quando são apresentados acordes maiores, menores e invertidos aos voluntários. Diferenças neurais também podem ser encontradas entre músicos que fazem papel de líder (Violino 1) e seguidor (Vio-

<sup>3</sup>O N1 inclui os subcomponentes N1a (pequeno e difuso, que ocorre em 75 ms pós-estímulo), N1b (fronto-central em 100 ms) e N1c (temporal em 150 ms)

<sup>4</sup>O P3 inclui os subcomponentes P3a, relacionado a resposta inconsciente a um estímulo, e o P3b, relacionado a detecção consciente do estímulo raro no paradigma *oddball* (POLICH, 2009)

lino 2) durante a execução de duetos (BALARDIN et al., 2017; VANZELLA et al., 2019), sendo que o seguidor apresenta maior ativação nas regiões temporo-parietais e somatomotoras comparando a condição em que o músico executava a peça solo com a condição em que o mesmo executava o dueto, em experimento utilizando fNIRS. Ademais, a estratégia de estudo empregada pelos músicos também pode ser um fator discriminante (SEPPÄNEN; BRATTICO; TERVANIEMI, 2007). Neste caso, é analisado o processamento auditivo de músicos que dão maior ênfase na teoria musical e leitura de partituras, com músicos que preferem improvisação, tocar de ouvido, e ensaiar ouvindo gravações, resultando em diferenças no processamento do contorno e intervalos melódicos entre os grupos, evidenciadas pelo MMN. Isso sugere que a plasticidade cerebral é sensível ao instrumento, gênero e às estratégias de treinamento de cada músico, influenciando na forma como eles processam a música.

Músicos amadores apresentam melhor discriminação de tons quando comparados com não-músicos (TERVANIEMI et al., 2006), como é evidenciado pela presença do componente P3a apenas no grupo de músicos, além deles apresentarem maior amplitude no componente MMN quando comparados com não-músicos. Isso demonstra que apesar da falta de um treinamento formal em teoria musical os músicos amadores apresentam mecanismos corticais mais avançados para codificar e categorizar a informação sonora quando comparados com não-músicos, especialmente para estímulos relacionados ao gênero musical que atuam, embora essa vantagem não possa ser generalizada para estímulos sonoros mais complexos (TERVANIEMI et al., 2006). Comparando músicos profissionais com amadores, resultados da literatura demonstram que profissionais exibem padrões de ativação emocional diferente e/ou mais intensos do que amadores, observados pela maior presença da banda teta nas regiões frontais, ressaltando a influência da experiência musical no processamento emocional durante a escuta musical (MIKUTTA et al., 2014). Sabe-se também que o contexto influencia na ativação cerebral durante a classificação da emoção sentida durante a escuta musical (MARKOVIC; KÜHNIS; JÄNCKE, 2017). Neste caso, foram analisadas as bandas de frequência dos sinais de EEG de voluntários registrados durante duas condições que eram apresentadas em diferentes ordens, sendo essas: (i) a escuta musical; e (ii) a escuta musical e avaliação simultânea dos níveis de valência e excitabilidade. Resultados demonstram aumento em todas as bandas de frequência durante a escuta musical, apenas quando não era precedida da condição de escuta e avaliação simultânea. Ou seja, se a primeira condição apresentada ao voluntário era a de escuta e avaliação simultânea, na sequência, os voluntários continuavam realizando mentalmente uma avaliação, muito embora eles deveriam apenas escutar a música. Existem, inclusive, indícios de que a catego-

rização de acordes maiores e menores ocorrem mesmo na ausência de treinamento musical, sendo que tanto músicos quanto não-músicos apresentam o componente N1 para esse tipo de estímulo, embora apenas os músicos apresentem o componente MMN para acordes menores e invertidos (VIRTALA et al., 2014). Além disso, as redes neurais de músicos e não-músicos diferem não somente durante a escuta musical, mas também durante o repouso (KLEIN et al., 2016), considerando que a atividade neuronal produzida durante o repouso reflete uma atividade funcional significativa, que é moldada pelo treinamento repetitivo. Assim, é demonstrado que regiões cerebrais que são ativadas de maneira síncrona durante a performance musical também são ativadas durante o repouso.

Angulo-Perkins et al. (2014) apresentaram resultados interessantes ao analisar os sinais de fMRI de um conjunto de 57 voluntários músicos e não-músicos em resposta a apresentação de diversos tipos de estímulos como sons de piano, frases vocalizadas, vocalizações de macacos e sons não-vocais como buzinas de carro. Eles encontraram que regiões corticais são ativadas por sons de categorias diferentes, ou seja, existem regiões funcionais específicas para cada tipo de estímulo apresentado. Além disso, foi encontrado maior ativação no *planum polare* em músicos, comparado com não-músicos, além de demonstrar que a prática musical modula a atividade gerada pelos estímulos musicais e frases vocalizadas no *planum temporale* direito.

Outros estudos propõem prever a musicalidade por meio das ativações cerebrais durante tarefas auditivas. Ao apresentar intervalos de quinta justa e tritonos para músicos profissionais e não-músicos, resultados demonstram uma acurácia de 95% na diferenciação dos dois grupos de voluntários, utilizando o valor médio das épocas registradas no eletrodo Cz em um grupo de 6 músicos e 6 não-músicos (LIANG et al., 2011). Em outro estudo análogo e mais recente, utilizando uma abordagem de análise multivariada para dados de fMRI obtidos durante a escuta de três peças musicais, foi apresentada uma acurácia de 77% de discriminação dos voluntários (18 músicos e 18 não-músicos) (SAARI et al., 2018). Semelhantemente, no contexto de análise multivariada para dados de EEG, encontramos, como será detalhado no próximo capítulo desta tese, uma acurácia de discriminação dos voluntários (13 músicos e 13 não-músicos) que varia de 62% (limiar inferior) a 92% (limiar superior) durante a escuta musical (RIBEIRO; THOMAZ, 2019).

Uma questão importante a ser levada em consideração é se essas diferenças encontradas entre músicos e não-músicos são unicamente resultado do treinamento musical ou se existe algum tipo de diferença pré-existente nas habilidades auditivas e motoras dos músicos que permite a eles terem um melhor desenvolvimento ao receberem o treinamento musical (ZATORRE;

CHEN; PENHUNE, 2007; BROWN; ZATORRE; PENHUNE, 2015; EGGERMONT, 2014; SCHLAUG, 2001). Isto é, “Os músicos nascem com alguma pré-disposição para a música?” (ZATORRE; CHEN; PENHUNE, 2007; BARRETT et al., 2013). Para responder essa pergunta, Hyde et al. (2009) realizaram um experimento longitudinal no qual crianças de aproximadamente 6 anos que não possuíam nenhum conhecimento musical foram submetidas a um exame de Imagem por Ressonância Magnética (do inglês, *Magnetic Resonance Imaging* - MRI) e a testes comportamentais específicos. Posteriormente, parte das crianças começaram a fazer aulas de piano, formando um grupo de crianças com treinamento musical, e as demais formaram o outro grupo de controle. Após 15 meses todas estas crianças foram submetidas aos mesmos exames. Como resultado, foram encontradas diferenças nas áreas motoras e auditivas entre os dois grupos, inerentes à prática musical. Schlaug et al. (2005) analisaram crianças de 5 a 7 anos durante 12 meses de treinamento musical. Diferentemente do grupo de controle, as crianças que estudaram música durante este tempo apresentaram aumento na ativação no lobo temporal bilateralmente e no giro temporal superior durante tarefas de discriminação rítmica e melódica. Embora esses resultados não possam eliminar o fator de pré-disposição para a música, os trabalhos de Hyde et al. (2009) e Schlaug et al. (2005) demonstram que o cérebro se altera como resultado do treinamento musical regular (EGGERMONT, 2014). Ou seja, durante a infância o cérebro é mais plástico, então os mecanismos neuroplásticos, que modificam conexões sinápticas ou que geram o crescimento de processos neurais, refletem as entradas sensoriais que são experienciadas durante a prática da música (TRAINOR; DESJARDINS; ROCKEL, 1999; PANTEV et al., 2001; NAGER et al., 2003; SEPPÄNEN; BRATTICO; TERVANIEMI, 2007; BARRETT et al., 2013).

Ainda neste contexto de investigar diferenças cognitivas entre músicos e não-músicos, é importante ressaltar que a grande maioria dos estudos de ERP é baseada em análises estatísticas massivamente univariadas. Ou seja, a maioria desses estudos explora as respostas do processamento neural localmente, em regiões em que se espera a ocorrência de um potencial evocado decorrente de um determinado estímulo. Dessa forma, não é levada em consideração a correlação entre as atividades neurais com outras áreas cognitivas. Portanto, vale destacar aqui os trabalhos realizados por Saari et al. (2018) e Ribeiro e Thomaz (2019), que são detalhados no próximo capítulo, que fizeram uso da análise estatística multivariada para predição de treinamento musical. Essa é uma abordagem válida porque a prática musical, como já visto, requer o uso paralelo de diversas áreas de processamento cognitivo. Em outras palavras, é necessário considerar as atividades neurais de maneira holística, observando as relações com outras áreas

cognitivas para, assim, permitir a compreensão da complexidade da informação resultante dos estímulos apresentados aos voluntários e as inter-relações destas áreas distintas.

### 3.3 A MÚSICA E O ENVELHECIMENTO

A interação entre o desenvolvimento cerebral e a aquisição de habilidades, especialmente no período da infância e adolescência, fornece provas de que a performance musical pode promover a geração e/ou retenção de tecidos corticais funcionalmente relevantes para a sua prática (SLUMING et al., 2002). Como dito anteriormente, tocar um instrumento musical requer não apenas o uso do processamento auditivo, mas também do motor, do visual e de processos relacionados à memória. Isso diferencia a música de outros modelos sobre plasticidade neural, devido a sua complexidade intrínseca (MORENO; BIDELMAN, 2014).

Os estudos descritos nas duas seções anteriores fornecem evidências científicas de que o treinamento musical está associado com mudanças neuroplásticas no cérebro, sendo a plasticidade neural definida como a habilidade do cérebro de se modificar ou alterar em função da demanda do ambiente (PASCUAL-LEONE et al., 2005). Esta prática a longo prazo pode facilitar as inter-relações funcionais de distintas áreas do cérebro de tal maneira que as suas características eletrofisiológicas podem ser detectadas até mesmo durante o repouso (*resting state*) (KLEIN et al., 2016). Sendo assim, os músicos formam um grupo importante para se estudar os efeitos da plasticidade cerebral a longo prazo (BENGTSSON et al., 2005), não apenas a nível cortical, mas também em áreas subcorticais (MORENO; BIDELMAN, 2014). Tendo em vista esses benefícios neuroplásticos gerados pela prática musical, será que os mesmos se estendem a idosos, que são mais suscetíveis ao declínio cognitivo?

Atualmente existe uma proporção crescente de idosos no mundo (UNITED NATIONS, 2015, 2017). A nível global, pessoas com idade maior ou igual a 65 anos são o segmento que cresce mais rapidamente da população (HORNBY-TURNER; PEEL; HUBBARD, 2017). Tendo isso em vista, é importante tomar medidas que possibilitem um envelhecimento saudável, de tal forma que gere oportunidades para melhorar e/ou preservar a saúde física, mental e social, assim como a independência e qualidade de vida desses indivíduos (VINCIGUERRA, 2017).

Problemas de audição em idosos normalmente se referem a dois tipos de patologias: (i) disfunção no ouvido interno (i.e. na cóclea) que resulta em perda auditiva, primariamente para altas frequências e perda mais moderada em baixas frequências; e (ii) processamento auditivo central anormal nas vias auditivas ascendentes e nas áreas auditivas primárias e associativas.

A perda auditiva relacionada ao envelhecimento resulta em sinais ruidosos sendo entregues ao sistema nervoso central, que interfere no processamento auditivo central (ALAIN et al., 2014).

Idosos, incluindo aqueles com audição normal, normalmente apresentam dificuldades em entender a fala na presença de outros sons concorrentes como, por exemplo, o som da TV, de música ou de outras pessoas falando. A compreensão da linguagem depende de como o cérebro utiliza a informação acústica que chega, sugerindo que processos neurais centrais, em oposição a mecanismos periféricos (e.g. mudanças na cóclea), tem um papel crítico no processamento das informações auditivas (ALAIN et al., 2014).

O envelhecimento pode também causar um lento enfraquecimento natural em algumas funções cognitivas como memória, velocidade de processamento da informação, aprendizado visual, habilidade para resolução de problemas e funções executivas. Essas disfunções são comumente associadas, no entanto, com mudanças comportamentais, depressão e ansiedade, que levam a piora na qualidade de vida, causando impactos negativos em várias funções, incluindo habilidades sociais e atividades diárias (VINCIGUERRA, 2017; HORNBY-TURNER; PEEL; HUBBARD, 2017). Nesse contexto, diversas pesquisas buscam meios para atrasar o declínio cognitivo com as mais diversas abordagens, afinal, pesquisadores têm demonstrado que o cérebro parece manter um nível impressionante de plasticidade ao longo da vida, até mesmo em idades avançadas (BHERER, 2015).

Quatro fatores possuem um papel crítico na forma como nos adaptamos ao envelhecimento: (i) interação social; (ii) nutrição; (iii) estimulação cognitiva; e (iv) atividade física (BHERER, 2015). Estar engajado com atividades físicas e intelectuais pode mitigar o declínio das funções cognitivas relacionado com o envelhecimento (ALAIN et al., 2014) de tal forma que a estimulação cognitiva e programas de atividades físicas podem aumentar momentaneamente a cognição (BHERER, 2015) ou atenuar (*offset*) o declínio cognitivo (BIDELMAN; ALAIN, 2015). Sabe-se também que um alto nível de educação está altamente associado com um positivo envelhecimento saudável (HORNBY-TURNER; PEEL; HUBBARD, 2017). Atividades musicais e dança são também alvo de estudos devido ao seu potencial para induzir plasticidade.

Comparada com outras atividades, a habilidade musical permanece relativamente preservada em estados avançados de demência e serve como meio de comunicação entre pacientes e cuidadores. Sabe-se que a música pode ser efetiva para reduzir agitação, ansiedade e depressão em pacientes com demência, mas é importante saber que a memória musical pode ser afetada diferentemente entre os diversos tipos de demência como o Alzheimer, a Demência Semântica ou a Demência Frontotemporal (BAIRD; SAMSON, 2015).

A importância de se estudar e analisar os impactos da prática musical no envelhecimento se dá devido ao fato de haver um crescente número de evidências científicas demonstrando que o estilo de vida proporciona um impacto significativo no envelhecimento saudável (ZENDEL; ALAIN, 2012). Essas evidências apontam que as mudanças geradas no cérebro por meio da plasticidade neural em função da prática musical podem ser convertidas em processamentos neurais relacionados a outros sons fora do escopo da música (e.g. linguagem) (MORENO; BIDELOMAN, 2014), além de poder gerar atraso ou atenuação do declínio na percepção e cognição auditiva (ALAIN et al., 2014). Assim, muitos estudos apontam que o treinamento musical intenso a longo prazo é capaz de alterar os efeitos do envelhecimento no processamento auditivo central durante a vida. Como dito anteriormente, a música atua facilitando a recuperação cognitiva, emocional e neural, aumentando as funções executivas, melhorando o humor e a qualidade de vida em idosos (SARKAMO, 2017; ALTENMÜLLER; SCHLAUG, 2015).

A abordagem comumente utilizada para se analisar o processamento cognitivo durante o envelhecimento envolve a seleção de dois grupos distintos, um possuindo voluntários jovens e o outro idosos. Desta forma, é possível analisar e inferir quais são as diferenças e os impactos do envelhecimento no processamento cognitivo. Funcionalmente, estudos utilizando EEG têm encontrado diferenças importantes com relação a músicos jovens e idosos. Por exemplo, investigando as respostas pré-atencionais, que apresentam um processamento rápido e inconsciente, é possível encontrar diferenças entre idosos músicos e não-músicos (SCHLAUG, 2015; O'BRIEN; NIKJEH; LISTER, 2015), sendo que conforme se aumenta a idade, observa-se uma diminuição na amplitude dos componentes de ERP analisados e um aumento na latência do pico do sinal de EEG, tanto para o MMN (KOK, 2000) quanto no complexo P1-N1-P2 (O'BRIEN; NIKJEH; LISTER, 2015). O mesmo efeito pode ser observado em repostas atencionais, que apresentam um processamento mais lento e consciente, como nos componentes P3 e P4. Sendo assim, os paradigmas do tipo *oddball*, no qual é necessário uma reação (e.g. apertar um botão, realizar um determinado movimento, etc.), também apresentam distinções quanto à amplitude e latência dos componentes de ERP entre idosos e jovens (KOK, 2000; HO et al., 2012; BIDELOMAN; ALAIN, 2015). Analisando-se os componentes de ERP no domínio da frequência, é possível encontrar diferenças nas ativações corticais em regiões distintas do córtex. Resultados sugerem que as respostas em diferentes bandas de frequência são moduladas tanto pelos processos cognitivos e sensoriais, quanto pela idade (HO et al., 2012; DUSHANOVA; CHRISTOV, 2014). Evidências sugerem ainda que para uma pessoa com maior idade executar uma tarefa com tanta eficiência quanto uma pessoa mais nova, uma importante diferença no padrão

de ativação cortical ocorre, havendo uma maior ativação cortical nas regiões frontais do cérebro (HO et al., 2012; DUSHANOVA; CHRISTOV, 2014; O'BRIEN; NIKJEH; LISTER, 2015), servindo como um meio de adaptação ao declínio cognitivo.

Todos estes resultados demonstram que pessoas idosas que tiveram uma prática musical a longo prazo apresentam um desempenho neural mais rápido em comparação com aqueles que não a tiveram (SARKAMO, 2017), evidenciando que a música é um forte estimulante para mudanças plásticas no cérebro, inclusive em adultos (MÜNTE; ALTENMÜLLER; JÄNCKE, 2002; ALTENMÜLLER; SCHLAUG, 2015; SCHLAUG, 2015).

No entanto, verifica-se que todos esses trabalhos citados utilizaram-se de estímulos artificiais para a realização dos experimentos, acarretando nas questões já descritas anteriormente. Adicionalmente, parte desses trabalhos compararam amostras de idosos com jovens, sem levar em consideração o treinamento musical. Outra parte desses trabalhos realizou a comparação de amostras de idosos músicos com idosos não-músicos. A única exceção é o trabalho de O'Brien, Nikjeh e Lister (2015), que fez o uso dos quatro grupos de interesse para a sua análise.

Estes estudos evidenciam a importante presença de uma Reserva Cognitiva gerada pelas ricas experiências vividas, na qual sujeitos com maior reserva cognitiva possivelmente apresentam efeitos menos severos ao processo de envelhecimento, sendo que o contínuo engajamento em atividades que estimulem o cérebro são capazes de manter ou aumentar estas reservas (HABECK et al., 2004; ZENDEL; ALAIN, 2012; ALAIN et al., 2014). Consequentemente, é esperado que a demanda neural requerida pela prática musical possa potencialmente reduzir o declínio cognitivo relacionado ao envelhecimento, e, embora muito interessante, tal tema encontra-se fora do foco desta tese.

### 3.4 RESERVA COGNITIVA

Conforme o ser humano envelhece ocorrem mudanças no estilo de vida que podem incluir a redução na mobilidade, menos oportunidades de socialização e participações menos frequentes de atividades que estimulem a cognição. Esse ponto deve ser levado em consideração, pois hoje se sabe que o estilo de vida influencia profundamente as estruturas e funções cerebrais (BARTOLOTTI; LAZAROV, 2016). Atividades físicas, interações sociais e participações em atividades que requerem demandas cognitivas como a leitura, escrita ou tocar um instrumento musical estão associadas com um menor risco para doenças neurodegenerativas (e.g. Doença de Alzheimer).

Uma interpretação para isso é que essas demandas cognitivas ao longo da vida constroem uma “reserva” que permite ao cérebro lidar melhor com os efeitos do declínio cognitivo relacionados ao envelhecimento ou a danos cerebrais. Assim, para que esses déficits se tornem aparentes, é necessário que estes estados estejam mais avançados (BARTOLOTTI; LAZAROV, 2016). Essa reserva pode existir de duas formas: Reserva Cerebral ou Reserva Cognitiva.

Reserva Cerebral se refere a diferenças quantitativas no cérebro dos indivíduos que permitem que algumas pessoas lidem melhor com danos cerebrais (STERN, 2009), ou seja, está relacionada com o tamanho do cérebro, quantidade de neurônios, entre outros fatores (Modelo passivo de reserva). Em contrapartida, a Reserva Cognitiva assume que as diferenças individuais nos processos cognitivos ou nas redes neurais subjacentes ao desempenho da tarefa permitem que algumas pessoas lidem melhor do que outras com danos cerebrais (ou efeitos do envelhecimento) (STERN, 2009), ou seja, o cérebro ativamente busca lidar com danos cerebrais usando processos cognitivos pré-existentes ou por meio de processos compensatórios (Modelo ativo de reserva). Embora sejam interdependentes em alguns aspectos, Reserva Cerebral e Cognitiva se sobrepõem e contribuem para a compreensão das diferenças entre indivíduos ao lidarem com danos cerebrais (STEFFENER; STERN, 2012).

Stern (2009) propõe que a Reserva Cognitiva seja subdividida ainda em Reserva Neural e Compensação Neural. A primeira se refere a variações entre indivíduos no processamento cognitivo em relação a eficiência e capacidade. Ou seja, quando há um aumento na demanda da tarefa, pessoas com maior eficiência necessitam de um menor aumento na atividade cerebral para processar essa demanda neural. Conforme se aumenta a demanda, atinge-se um limite máximo na rede neural, sendo esse um índice da capacidade neural (STEFFENER; STERN, 2012). A segunda se refere a variações entre indivíduos em compensar alterações no processamento cognitivo em resposta a alguma patologia cerebral. Essa compensação pode ajudar a manter ou melhorar a performance cognitiva, fazendo uso de estruturas e redes neurais que normalmente não são utilizadas por indivíduos normais (estratégias cognitivas) (STEFFENER; STERN, 2012).

A teoria da reserva cognitiva foi desenvolvida em pesquisas sobre envelhecimento e demência (BARTOLOTTI; LAZAROV, 2016). Em doenças neurodegenerativas, a reserva cognitiva é considerada como uma proteção contra manifestações clínicas, sugerindo que uma pessoa pode aumentar a sua reserva cognitiva com atividades ocupacionais e educacionais. Ou melhor, um indivíduo que experimentou ao longo de sua vida um determinado nível de complexidade cognitiva pode ter gerado conexões neurais alternativas que podem beneficiá-lo em sua vida

(BARTOLOTTI; LAZAROV, 2016), reduzindo o risco de desenvolver demências e apresentando menor taxa de declínio da memória, comum no processo de envelhecimento (STERN, 2009). Além disso, sabe-se que o período de desenvolvimento na infância é crucial para a construção de reservas cognitivas, porém as circunstâncias da vida e as atividades empregadas continuam a influenciar no desenvolvimento dessas reservas (STEFFENER; STERN, 2012).

Músicos podem ser um excelente modelo para estudar reserva cognitiva, especialmente em termos de declínio relacionado ao envelhecimento, relacionada ao processamento da informação auditiva (ALAIN et al., 2014). Por exemplo, sabe-se que quando os sistemas auditivos centrais e periféricos não conseguem processar efetivamente a linguagem, regiões pré-frontais podem ser recrutadas para compensação, assim facilitando a compreensão da linguagem em ambientes ruidosos (ALAIN et al., 2014). Músicos, diferente de não-músicos, recrutam pequenas áreas nas regiões motoras do cérebro, sugerindo que o treinamento musical a longo prazo pode resultar em uso mais eficiente dos recursos neurais disponíveis (CHEN; PENHUNE; ZATORRE, 2008). O treinamento musical pode criar uma reserva cognitiva, na qual permite que músicos mais velhos mantenham as habilidades auditivas cognitivas (i.e. escuta musical) em idades avançadas (ALAIN et al., 2014). De novo, embora muito interessante, este tema encontra-se fora do foco desta tese.

### 3.5 DESCRITORES SONOROS

Como ouvintes, buscamos compreender e desvendar as estruturas musicais por meio dos seus padrões, suas repetições e suas transições. Um músico consegue compreender esses elementos e estruturas, tendo como base o seu conhecimento de teoria musical ou a sua prática e experiência. No entanto, mesmo um leigo consegue perceber intuitivamente as diferenças nos contornos melódicos, nas mudanças de tonalidade, no ritmo e nos instrumentos empregados em uma determinada música, tendo como referência informações que estão armazenadas em sua memória, relacionando estruturas que são familiares a ele. Assim, nosso cérebro é capaz de extrair, processar e interpretar esses elementos musicais.

A sucessão de sons produzidos durante a execução de uma determinada música apresenta um conjunto de padrões sonoros, especificados normalmente nas partituras musicais ou dependentes da interpretação do intérprete (músico), que possuem propriedades acústicas como a altura tonal, o timbre, a duração e a intensidade dos sons.

No contexto de processamento de sinais, é possível extrair e analisar informações relevantes dos padrões sonoros em áudios musicais por meio da extração de características acústicas, reduzindo a quantidade de dados que precisam ser processados, obtendo uma descrição compacta e expressiva do conjunto de dados originais. Nesse contexto, existem muitos descritores sonoros que podem ser utilizados para diferentes propósitos como a classificação de gêneros musicais, similaridade musical, detecção de emoções, entre outros. Estas técnicas de análise de sinais de áudio deram origem a área de pesquisa científica multidisciplinar conhecida como Recuperação de Informação Musical (do inglês, *Music Information Retrieval* - MIR).

As características acústicas que podem ser extraídas dos áudios são descritores sonoros que representam propriedades específicas destes sinais (MITROVIĆ; ZEPPELZAUER; BREITENEDER, 2010). Essas características descrevem o som de forma quantitativa ou qualitativa, podendo ser categorizadas em diferentes níveis de abstração. Características de baixo nível (do inglês, *Short-term features*) são tipicamente computadas diretamente dos sinais de áudio, descrevendo a forma de onda estatisticamente (KNESS; SCHEDL, 2016); no entanto, essas características não apresentam necessariamente um significado musicológico ou perceptual (relativo à percepção humana) (LERCH, 2012). Características de alto nível (do inglês, *Long-term features*) descrevem propriedades do sinal de áudio que são semanticamente significativas (e.g. andamento, acordes, propriedades melódicas) (LERCH, 2012), descrevendo estes sinais em termos semelhantes aos percebidos pelos humanos (KNESS; SCHEDL, 2016), e podem ser construídas a partir das características de baixo nível.

As características acústicas podem ser extraídas diretamente no domínio do tempo, que buscam analisar como o sinal se comporta ao longo do tempo, avaliando os aspectos do desenvolvimento temporal do sinal de áudio. Também podem ser extraídas no domínio da frequência por meio da Transformada Rápida de Fourier (do inglês, *Fast-Fourier Transform* - FFT), aplicada a janelas de tempo de alguns milissegundos, correspondendo a Transformada de Fourier de Tempo Curto (do inglês, *Short-time Fourier Transform* - STFT), avaliando a relação de amplitude entre os harmônicos formados (KNESS; SCHEDL, 2016; LERCH, 2012).

A obtenção dos descritores acústicos ou sonoros é uma importante etapa tanto para a classificação destes dados, quanto para a análise dos sinais cerebrais durante a escuta musical, principalmente em situações em que são apresentadas músicas completas para análise. O nosso cérebro é capaz de avaliar os aspectos dessa informação sonora, interpretando, reconhecendo, memorizando e processando toda essa informação. Pesquisas com EEG sobre como o cérebro processa a música utilizam tradicionalmente como abordagem a apresentação de estímulos ar-

tificiais aos sujeitos durante experimentos de escuta musical (POIKONEN et al., 2016). Como dito anteriormente, estes estímulos artificiais são construídos especificamente para os experimentos propostos, apresentando elementos isolados da música. Ou seja, eles apresentam apenas uma pequena parcela das múltiplas dimensões que constituem o que chamamos de música, demonstrando como o cérebro processa esses elementos isoladamente.

Quando ouvimos uma orquestra tocando, uma banda ou um grupo, diversos elementos da música são apresentados de forma paralela e não isoladamente, como acontecem com os estímulos artificiais. É necessário, portanto, encontrar formas para se analisar o processamento musical considerando todos os aspectos da música. Como uma alternativa para essa questão, estudos recentes baseados em EEG e fMRI têm utilizado músicas completas, ou pequenos trechos musicais (ALLURI et al., 2012, 2013; ABRAMS et al., 2013; POIKONEN et al., 2016; POIKONEN; TOIVIAINEN; TERVANIEMI, 2016; MARKOVIC; KÜHNIS; JÄNCKE, 2017; SAARI et al., 2018; FERREIRA; RIBEIRO; THOMAZ, 2019; RIBEIRO; THOMAZ, 2019) com o intuito de descrever as associações entre as mudanças dinâmicas das músicas com as ativações neurais que são registradas dos sujeitos durante a escuta musical.

Como proposto por Alluri et al. (2012) e Poikonen et al. (2016), é possível utilizar características acústicas extraídas dos áudios e identificar instantes capazes de gerar ativações neurais significativas durante a escuta musical, relacionadas com os ERPs. Para se analisar músicas completas é possível então utilizar descritores sonoros (ver Apêndice A), que são características acústicas de baixo nível, referências na literatura (LARTILLOT, 2014; ALLURI et al., 2012; POIKONEN et al., 2016), e que apresentam propriedades como:

- a) *Tonalidade*, que estima a quantidade de componentes tonais existentes no sinal em oposição a ruídos. A sensação de altura permite quantificar um som como grave ou agudo, estando diretamente relacionada com a frequência. Sons ruidosos apresentam espectros contínuos, enquanto sons tonais apresentam frequências específicas. Por exemplo, o ruído branco apresenta um espectro plano, com intensidade igual em diferentes frequências e com mínima tonalidade, enquanto uma onda senoidal pura apresenta alta tonalidade;
- b) *Intensidade*, que é uma medida física da magnitude do som. Em outras palavras, é a quantidade de energia que uma determinada fonte sonora produz;
- c) *Forma Espectral*, que são medidas de estatística descritiva e representam a dispersão do sinal no domínio da frequência, relacionada com a propriedade do timbre;

- d) *Brilho*, que caracteriza a distribuição espectral das frequências, apresentando se um sinal é dominado por frequências altas ou baixas;
- e) *Dissonância*, que é uma sensação de desarmonia, ou seja, a combinação de sons pouco ou nada harmônicos. É a percepção subjetiva de modulações rápidas de amplitude do som e frequências muito próximas.

### 3.6 CONSIDERAÇÕES COMPLEMENTARES

De acordo com os estudos descritos, há fortes evidências de que as diferenças existentes nos cérebros de músicos são mais um resultado da adaptação humana ao rigoroso treinamento musical (EGGERMONT, 2014), especialmente em períodos críticos do desenvolvimento cerebral, do que um resultado de propriedades inatas como uma pré-disposição para a música (HUTCHINSON et al., 2003). Evidências também apontam para a existência de fatores pré-existentes para a música (BROWN; ZATORRE; PENHUNE, 2015), porém entende-se que exista na verdade um equilíbrio entre esses dois fatores (predisposição e experiência) que juntos contribuem para a formação do músico. Existem expectativas de que essas diferenças geradas pela plasticidade neural podem servir como uma intervenção positiva para crianças com distúrbios no desenvolvimento, ou em adultos com doenças neurológicas (HYDE et al., 2009). Outros trabalhos sugerem que os métodos de ERP podem ser utilizados como ferramenta para seleção e recrutamento de músicos ou para avaliar musicalmente estudantes de música (VUUST et al., 2012), mas para que isso ocorra, é necessário que o desenvolvimento de metodologias que se utilizem do ERP garantam confiabilidade e sensibilidade a essa possível ferramenta.

Embora tenhamos destacado neste capítulo a importância da plasticidade neural, é importante ressaltar que muita plasticidade pode ser prejudicial (BARRETT et al., 2013), como é o caso da distonia focal, que é uma condição clínica que causa movimentos indesejados de músculos, podendo ser resultado de uma plasticidade mal adaptada. Sendo assim, as diferenças neurais entre músicos e não-músicos resultantes da prática musical só podem ser consideradas benéficas se tais diferenças resultem em um aprimoramento funcional.

Portanto, podemos observar a importância da música e como a sua prática causa impactos positivos em nossas vidas. Tendo isso em vista, desenvolver métodos que sejam capazes de demonstrar matematicamente que essa prática proporciona efeitos cognitivos significativos é de extrema importância.

## 4 PADRÕES COGNITIVOS EM ESCUTAS MUSICAIS

Neste capítulo é descrito um arcabouço computacional proposto nesta tese para abordar o problema de classificação de 2 grupos de amostras baseado em dados de sinais cerebrais, registrados por um equipamento de EEG e extraídos de voluntários músicos e não-músicos durante uma tarefa auditiva, para prever se uma determinada pessoa é um músico ou não (RIBEIRO; THOMAZ, 2019).

Como visto anteriormente, a prática musical é uma experiência intensa capaz de promover profundas mudanças no cérebro. Um músico está constantemente envolvido e exposto à música, mas existem diferenças intra-classes que estão relacionadas com as diferentes demandas neurais requeridas por cada instrumento musical (BANGERT; SCHLAUG, 2006; VUUST et al., 2012). Vimos, ainda, que músicos amadores também apresentam distinções em seus padrões de ativações cerebrais quando comparados com não-músicos (TERVANIEMI et al., 2006). Assim, temos uma grande variação no grupo ou classe de “músicos”. Isso torna este problema de classificação em algo não-trivial quando somado a natureza dos sinais utilizados para a classificação que, além de apresentarem alta dimensionalidade, normalmente sofrem com o problema de pouca quantidade de amostras. Podemos verificar essa afirmação pelos poucos trabalhos publicados buscando prever proficiência musical (LIANG et al., 2011; SAARI et al., 2018; RIBEIRO; THOMAZ, 2019) por meio de sinais cerebrais, sejam esses de EEG ou fMRI.

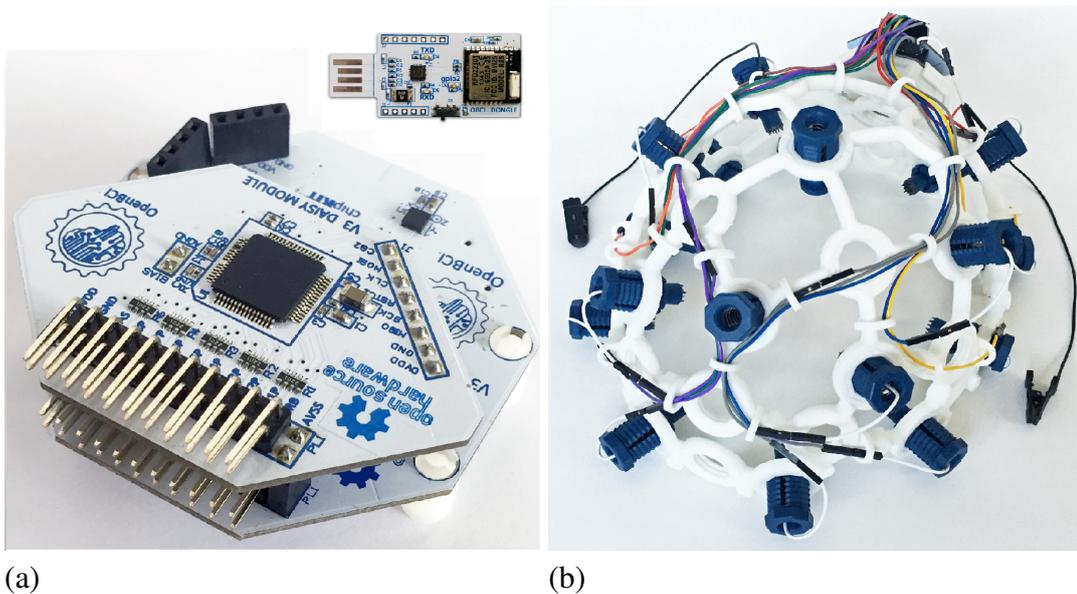
Assim, utilizando uma abordagem de estatística multivariada, que busca analisar os sinais de EEG como um todo, ou seja, de maneira holística, o arcabouço computacional proposto é capaz de promover uma melhor descrição dos padrões de ativações neurais, das regiões de transição e fronteiras de decisão que separam os grupos amostrados, indicando uma separação linear plausível entre estes grupos.

### 4.1 MATERIAIS

Para avaliação experimental deste arcabouço, foi utilizado o equipamento de EEG OpenBCI para a aquisição dos sinais elétricos (OPENBCI, 2017). Este equipamento possui uma frequência de amostragem de 125 Hz, uma resolução de 32 bits por canal e é composto por 16 eletrodos a seco de Cloreto de Prata (AgCl) posicionados no escalpo de acordo com o sistema 10-20 (MARCUSE; FIELDS; YOO, 2016), incluindo 2 eletrodos adicionais posicionados em cada

orelha como referência. São utilizados a placa Cyton (8 canais de eletrodos) e o módulo Daisy (8 canais de eletrodos sobressalentes) para realizar o registro, e o módulo USB para comunicação com o computador. Os eletrodos são fixados na estrutura denominada Cortex IV, sendo que a placa Cyton e o módulo Daisy se comunicam com o módulo USB via Bluetooth, evitando fios pendurados, trazendo maior praticidade e conforto para a realização dos experimentos, conforme apresentado na Figura 11.

Figura 11 – Materiais utilizados para aquisição dos sinais de EEG.



Fonte: Adaptado de OpenBCI, 2017.

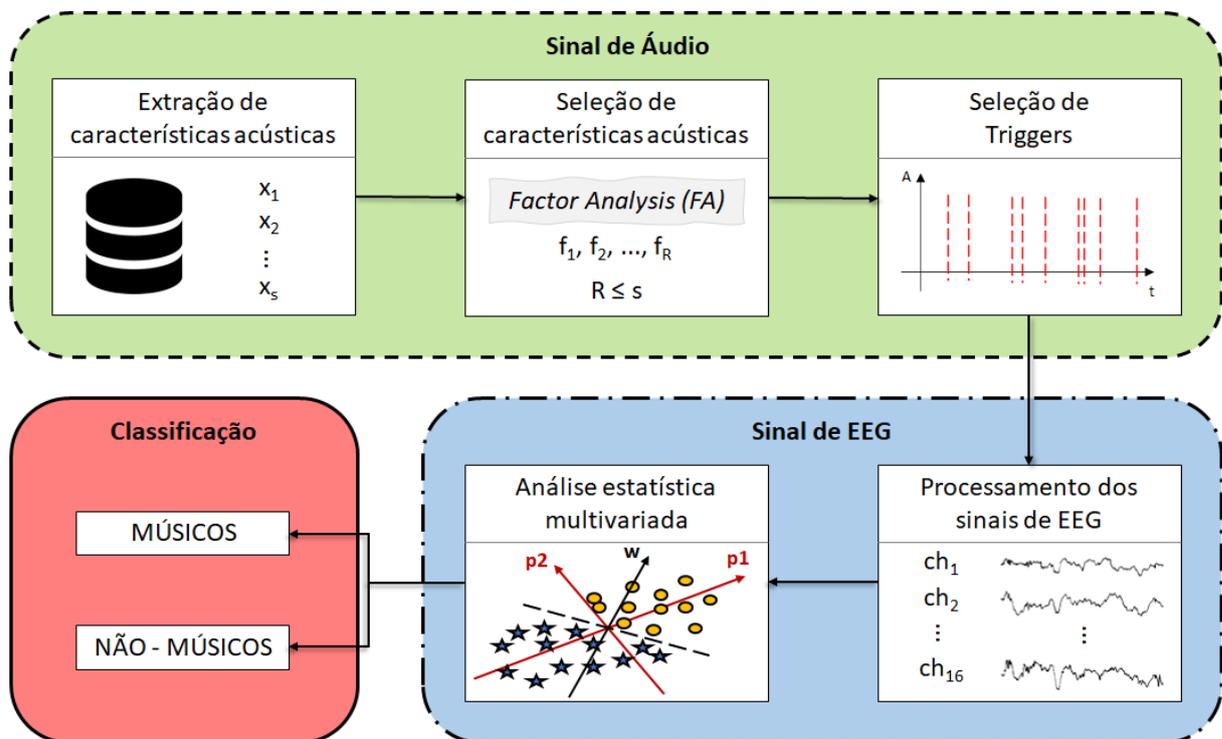
## 4.2 METODOLOGIA

Abordagens computacionais de aprendizado de máquina (*machine learning*) forneceram poderosas ferramentas para neurocientistas analisarem os dados neurais adquiridos em diversos experimentos nas últimas décadas (DAS; GIESBRECHT; ECKSTEIN, 2010). Neste trabalho são utilizadas ferramentas de reconhecimento de padrões para se analisar sinais de EEG, baseadas em estatística multivariada.

O arcabouço computacional proposto consiste nas seguintes 5 etapas: (I) Extração de características acústicas; (II) Seleção de características acústicas; (III) Seleção de Triggers; (IV) Processamento dos sinais de EEG; e (V) Análise estatística multivariada. As três primeiras etapas se referem ao sinal de áudio e as duas últimas ao sinal de EEG. Esta metodologia, ilustrada na Figura 12, foi construída tendo como base estudos multidisciplinares anteriores relacionados

ao processamento da música utilizando EEG (POIKONEN et al., 2016; POIKONEN; TOIVI-AINEN; TERVANIEMI, 2016) e a análise de dados em espaços de alta dimensão (THOMAZ et al., 2007; SATO et al., 2008; XAVIER et al., 2015; GREGORI; SANCHES; THOMAZ, 2017; ROCHA et al., 2014).

Figura 12 – Fluxograma das cinco principais etapas do arcabouço computacional proposto para analisar estatisticamente os sinais de EEG, sendo essas: (I) Extração de características acústicas; (II) Seleção de características acústicas; (III) Seleção de Triggers; (IV) Processamento dos sinais de EEG; e (V) Análise estatística multivariada.



Fonte: Ribeiro e Thomaz, 2018, 2019.

Para análise do sinal de áudio, tendo em vista a escolha de utilizar músicas completas como estímulos frente ao modelo tradicional de análises que se baseiam em estímulos artificiais, o uso da estatística multivariada, por meio da Análise Fatorial, busca encontrar relações entre as características acústicas utilizadas, a fim de obter uma melhor representação desses dados. Assim, a princípio, questiona-se a necessidade de utilizar todas as características disponíveis para a análise, mantendo apenas aquelas que são não-redundantes. Desta forma, promove-se a redução da dimensionalidade e do processamento dos dados.

Quanto ao sinal de EEG, vemos que a maioria dos estudos de ERP tem obtido resultados em um contexto local, baseados no conceito de análise estatística univariada, realizando um número limitado de ANOVAs (Análise de Variância) feitas em medidas pontuais tais como

a média ou amplitude dos picos dos sinais registrados pela técnica de ERP. Assim, computa-se a amplitude média nas épocas correspondentes aos fenômenos analisados (e.g. P3 com janelas de 300 a 600 ms pós-estímulo) e aplica-se, por exemplo, repetidas medidas de ANOVA com dois fatores: o tipo de estímulo e a localização do eletrodo (GROPPE; URBACH; KUTAS, 2011). Mais comum em experimentos de fMRI, mas que vem sendo implementado em experimentos de ERP, é o conceito de análise estatística massivamente univariada (*mass-univariate analysis*), definida como a análise de um grande número de variáveis dependentes medidas simultaneamente por meio da aplicação de técnicas de estatística univariada (e.g. Teste T). Assim, computa-se, por exemplo, um Teste T em cada ponto no tempo e eletrodo. A análise massivamente univariada permite observar com maior detalhamento espacial e temporal os efeitos dos ERPs (GROPPE; URBACH; KUTAS, 2011), apesar da necessidade de realizar correções para múltiplas comparações, que ocorrem quando a análise estatística envolve múltiplos testes estatísticos simultâneos, como é o caso. Além disso, essas abordagens costumam se basear em eletrodos individuais ou em um grupo selecionado de eletrodos em regiões específicas (YANG; ZHU; TIAN, 2018). Embora essas análises estatísticas univariadas e massivamente univariadas possam identificar localmente efeitos específicos do processamento musical, tais técnicas acabam ignorando a correlação espacial com outras áreas cognitivas. Ou seja, ainda que essas abordagens sejam matematicamente corretas, ignoram as possíveis dependências globais e inter-regionais do cérebro durante o processamento da música, provendo uma compreensão limitada da atividade neural durante tal tarefa cognitiva (RIBEIRO; THOMAZ, 2019). Como os dados de EEG incluem medidas simultâneas de um determinado número de eletrodos, acredita-se que seja mais apropriado investigar as relações entre todos os eletrodos por meio de métodos estatísticos multivariados para compreender a complexidade da informação contida nesses sinais.

Nesse contexto, é possível fazer uma análise global e holística dos sinais de EEG, considerando a atividade neural registrada em todos os eletrodos simultaneamente. Mais especificamente, são consideradas as interações entre regiões cerebrais com o propósito de medir, explicar e prever o grau de relacionamento entre as variáveis e seus impactos. Assim, a análise estatística multivariada lida com dados que incluem medidas simultâneas em várias variáveis. Compreender as relações entre tantas variáveis não é um problema simples, requerendo transformações algébricas distintas das análises univariadas (JOHNSON; WICHERN, 2007).

#### 4.2.1 Extração de características acústicas

Nesta primeira etapa, são extraídas características acústicas comumente utilizadas para análise de áudio na área de Recuperação de Informação Musical (MIR), por meio da MIRtoolbox (version 1.6.1) (LARTILLOT, 2014). Esta primeira etapa do processamento do sinal de áudio tem como finalidade resolver o problema da natureza dos estímulos neurais apresentados aos voluntários quando se pretende estudar como o cérebro processa a música. Lembremos, portanto, que tradicionalmente são utilizados estímulos artificiais para este tipo de análise, pois esses permitem um controle sobre o estímulo apresentado e a resposta neural esperada, baseados nos conceitos de ERP. Além disso, evita-se ter que trabalhar com o sinal contínuo de EEG.

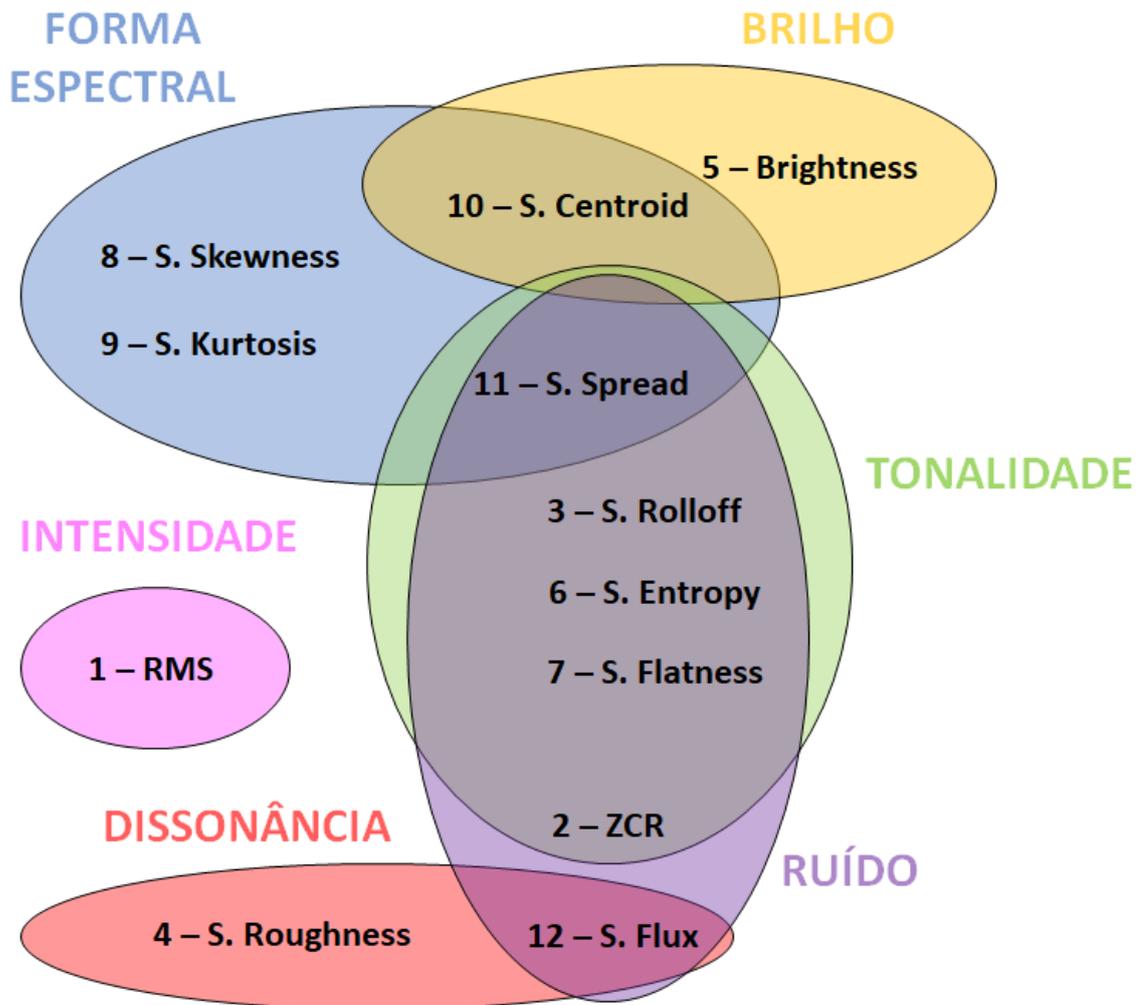
Porém, há a necessidade aqui de analisar os sinais neurais durante a escuta musical de músicas completas. Visando contornar este problema, Alluri et al. (2012) propuseram o uso de descritores acústicos para encontrar instantes nas músicas que fossem capazes de gerar respostas neurais significativas, utilizando fMRI. Posteriormente, Poikonen et al. (2016) aplicaram esta mesma análise para sinais de EEG, encontrando os componentes N1-P2 do potencial evocado para os descritores acústicos utilizados. Assim, em vez de analisar o sinal contínuo de EEG ao longo da música, analisa-se apenas estes instantes (denominados *triggers*) comprovadamente significativos. Dessa forma, este método permite que os voluntários sejam expostos a estímulos musicais completos, diferentemente daqueles estímulos artificiais utilizados tradicionalmente.

A extração das características acústicas é realizada com o objetivo de estudar a evolução de cada um desses descritores ao longo do tempo. Para isso, o sinal de áudio deve ser decomposto em pequenas janelas cujo tamanho comumente utilizado, de acordo com a literatura, é de 25 milissegundos (ms) (TZANETAKIS; COOK, 2002). No entanto, a MIRtoolbox utiliza como padrão janelas de 50 ms. Assim, para definir o tamanho da janela a ser utilizada é necessário compreender que esta etapa visa a identificação dos *triggers* no sinal de áudio. Em um estudo recente, Ferreira (2019) demonstrou que o uso de janelas de 50 ms de duração gera melhores resultados para a seleção de *triggers*. Portanto, nesta primeira etapa o sinal de áudio é decomposto em janelas de 50 ms de duração com 50% de sobreposição. Deste modo, similar a Poikonen et al. (2016) e Alluri et al. (2012), são utilizadas características acústicas de baixo nível para analisar os sinais de áudio, a fim de encontrar respostas neurais correspondentes aos potenciais evocados.

Foram selecionadas 12 características acústicas de referência na literatura (LARTILLOT, 2014), descritas no Apêndice A e ilustradas na Figura 13, sendo essas: (1) RMS; (2)

ZCR; (3) Spectral Rolloff; (4) Spectral Roughness; (5) Brightness; (6) Spectral Entropy; (7) Spectral Flatness; (8) Spectral Skewness; (9) Spectral Kurtosis; (10) Spectral Centroid; (11) Spectral Spread; e (12) Spectral Flux.

Figura 13 – Descrição das 12 características acústicas utilizadas nesta tese ordenadas de acordo com suas propriedades.



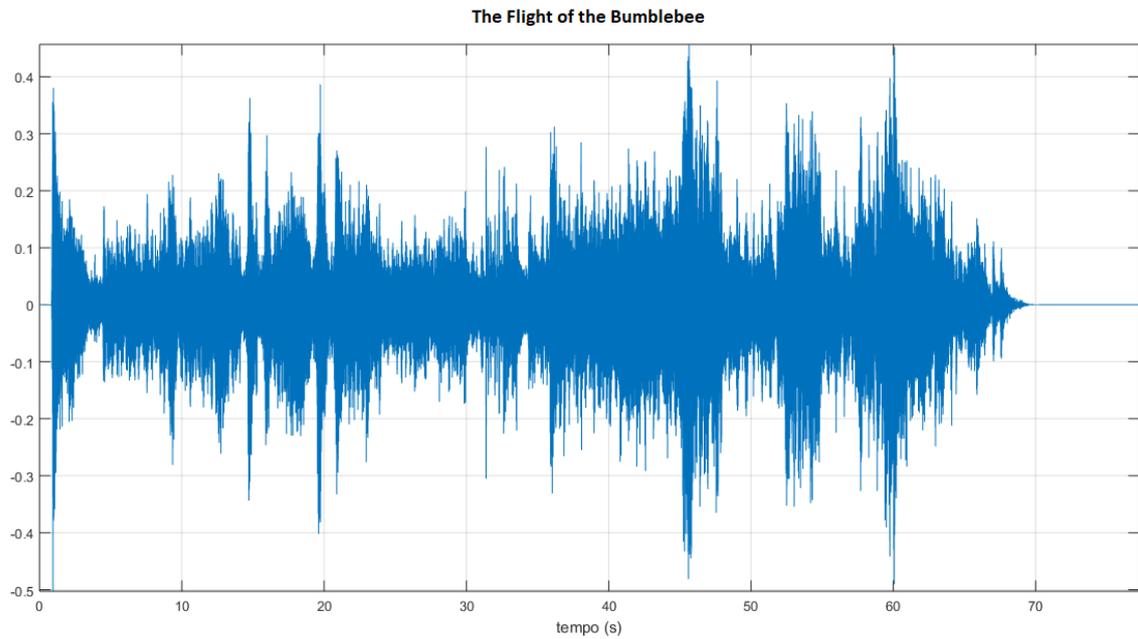
Fonte: Autor.

As Figuras 14 e 15 ilustram um exemplo de aplicação desta primeira etapa de extração de características apresentando, respectivamente, o sinal de áudio e a série temporal resultante da extração da característica acústica RMS para a música “*The Flight of the Bumblebee*” de Nikolai Rimsky-Korsakov, interpretada por Charles Gerhardt, James Galway e *The National Philharmonic Orchestra*<sup>1</sup>. É possível observar a evolução temporal da música analisada. Muito embora esse seja um trecho curto, essa música busca evocar o padrão de voo do besouro, sendo

<sup>1</sup>Disponível em: <https://open.spotify.com/track/7IpPpSisR1kR2E2ErOG0Ky?si=995zXEXIQJGjVe2KCjZU-Q>

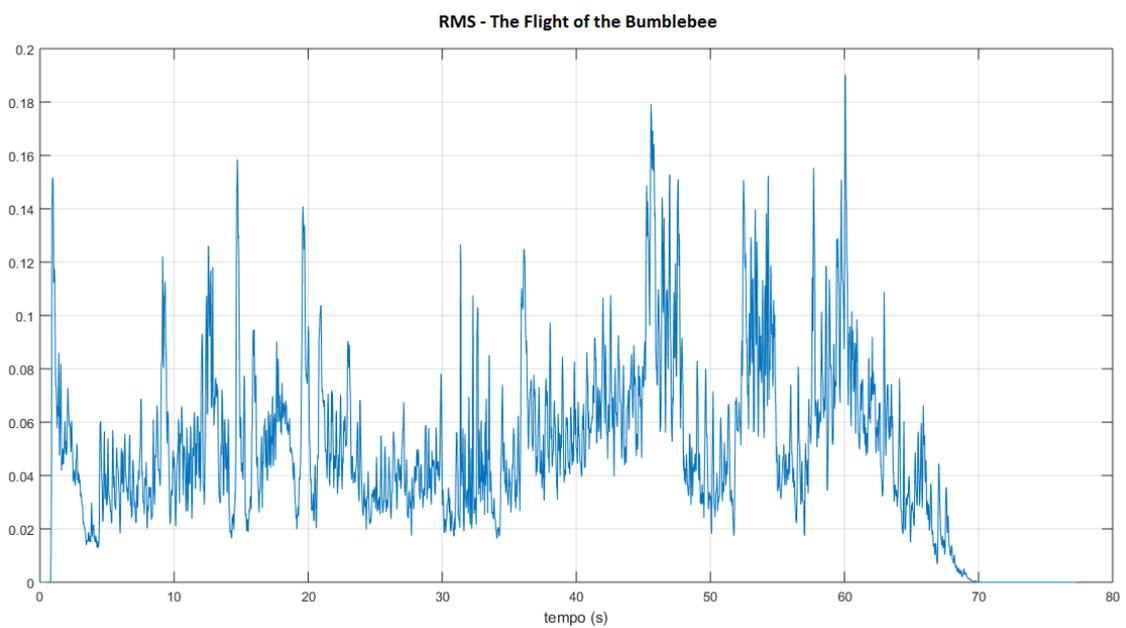
caótico e com mudanças muito rápidas. Assim, vemos a grande variação na energia do sinal apresentado pelo RMS e demonstrada na Figura 15.

Figura 14 – Sinal de áudio da música “*The Flight of the Bumblebee*” de Nikolai Rimsky-Korsakov.



Fonte: Autor.

Figura 15 – Sinal da característica acústica RMS extraída da música “*The Flight of the Bumblebee*” de Nikolai Rimsky-Korsakov.



Fonte: Autor.

#### 4.2.2 Seleção das características acústicas

Nesta segunda etapa, são avaliadas se todas as  $s = 12$  características acústicas extraídas do áudio são estatisticamente relevantes (*não-redundantes*), por meio de uma análise de agrupamento (*cluster*), revelando a correlação entre essas características para representar adequadamente os dados.

Para isso, utiliza-se a Análise Fatorial, ou *Factor Analysis* (FA) (JOHNSON; WICHERN, 2007), que é uma técnica conhecida de estatística multivariada, usada neste trabalho para descrever a associação entre os valores extraídos das janelas de cada característica acústica de modo não-supervisionado (RIBEIRO; THOMAZ, 2019), baseada na correlação entre essas variáveis. Objetiva-se reduzir a redundância dos dados extraídos do sinal de áudio, selecionando os  $R$  fatores mais significativos, sendo  $R \leq s$ , que explicam a inter-relação entre estas características acústicas.

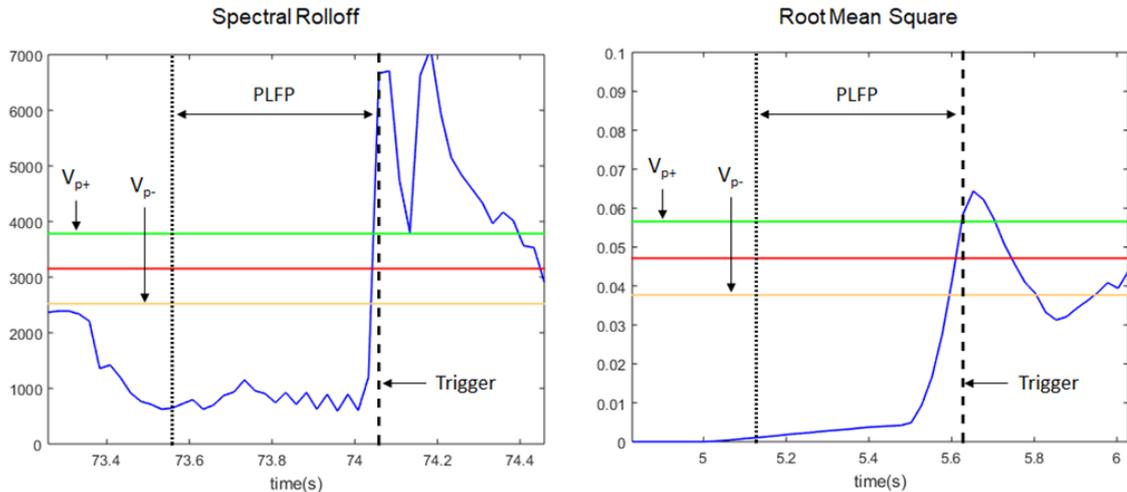
Para estimar os parâmetros deste modelo, é utilizado o método das componentes principais (JOHNSON; WICHERN, 2007), pois a sua solução de decomposição espectral simplifica a questão de quantos fatores devem ser retidos na FA (RIBEIRO; THOMAZ, 2019), convencionalmente escolhendo os fatores cujos autovalores sejam maiores do que 1 para determinar o número adequado de fatores utilizados na análise (JOHNSON; WICHERN, 2007). Assim, são extraídos os  $R$  fatores comuns rotacionados da FA (KAISER, 1958). Por fim é feita uma análise dos agrupamentos gerados entre as  $s$  características acústicas nos gráficos de dispersão das cargas fatoriais dos  $R$  fatores. Assim, elegem-se as  $sf$  (*selected features*) características acústicas com maior carga fatorial dentro de cada um dos agrupamentos identificados. A seção 5.1.2 apresenta uma descrição detalhada desta etapa.

#### 4.2.3 Seleção de triggers

Estímulos artificiais ocorrem em tempos específicos e pré-determinados de acordo com o protocolo de experimentos realizados, com a intenção de gerar respostas neurais resultantes. Esse comportamento, no entanto, não ocorre em músicas completas devido a sua complexidade. Portanto, é necessário identificar instantes nas músicas que sejam capazes de provocar uma resposta sensorial semelhante às encontradas nos estímulos artificiais. A estes instantes é dado o nome de *Triggers*, definidos como instantes nas séries temporais gerados pelas características

acústicas em que ocorrem um alto contraste e que são capazes de gerar os componentes N1-P2 do ERP (POIKONEN et al., 2016), como será detalhado a seguir.

Figura 16 – Ilustração da ocorrência de *triggers* para as características Spectral Rolloff e Root Mean Square (RMS) na música Dança Húngara No.5 de Johannes Brahms.



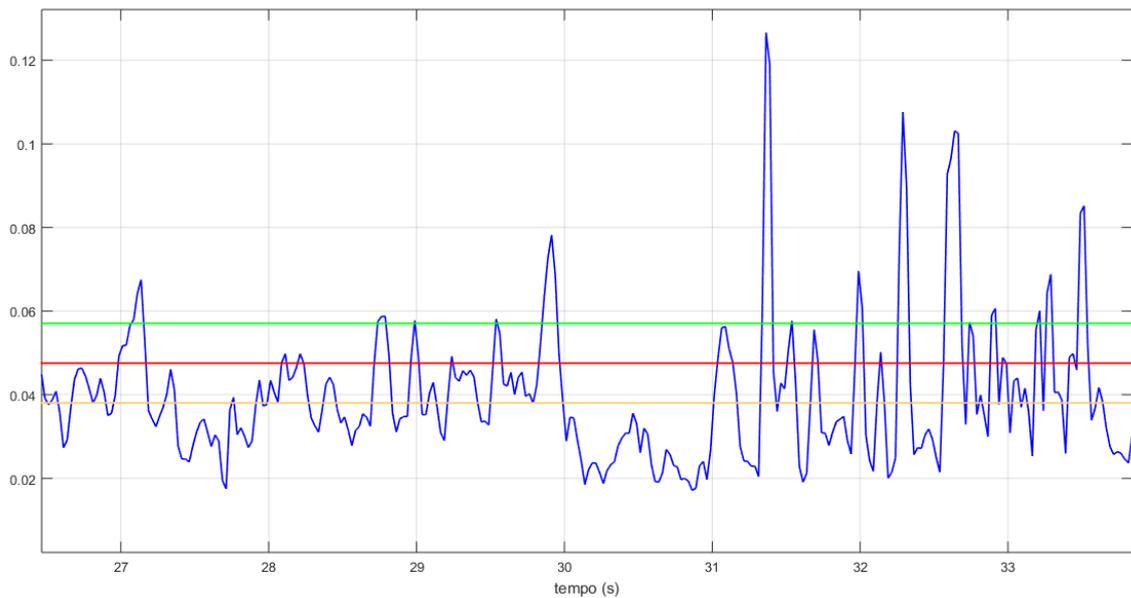
Fonte: Autor.

Foi adotado o método proposto por Poikonen et al. (2016) para identificar os *triggers*, sendo inicialmente necessária a definição dos limiares superiores e inferiores  $V_{p+}$  e  $V_{p-}$ , determinados como uma porcentagem  $q$  acima ou abaixo dos valores médios da característica acústica correspondente, tal que  $q = \pm 20\%$  (RIBEIRO; THOMAZ, 2019). Um *trigger* acontece quando o sinal permanece abaixo de  $V_{p-}$  por um intervalo de tempo mínimo denominado *Preceding Low-Feature Phase* (PLFP), definido aqui como sendo 500 ms para todas as características acústicas analisadas, seguido por uma fase ascendente em que o sinal atinge  $V_{p+}$  (POIKONEN et al., 2016), conforme demonstra a Figura 16. A intensidade do contraste acústico depende dos parâmetros definidos para a seleção dos *triggers* (tamanho do PLFP,  $V_{p+}$  e  $V_{p-}$ ). Estes parâmetros podem ser alterados baseados em um compromisso entre o número de *triggers* identificados e quão alto o contraste acústico pode ser definido, ou seja, quanto maior o contraste acústico, menor o número de *triggers* identificados (RIBEIRO; THOMAZ, 2019).

No entanto, vemos por esse método que o mesmo depende exclusivamente da ocorrência dos contrastes nas séries temporais resultantes das características acústicas extraídas do áudio (ilustrados na Figura 16). Porém, dependendo da construção da música é possível que esse contraste não aconteça. Um exemplo é o caso da música ‘*The Flight of the Bumblebee*’ de Nikolai Rimsky-Korsakov, apresentada na Seção 4.2.1, que apresenta uma progressão com rápidas mu-

danças de notas. Se observarmos mais de perto o sinal extraído da característica RMS desta música (Figura 17), podemos ver claramente a não-ocorrência de *triggers* para os parâmetros aplicados. Como dito anteriormente, é possível alterar os parâmetros para identificação desses instantes, mas pode haver casos em que não sejam encontrados *triggers*, invalidando o uso deste método para esse tipo de música.

Figura 17 – Trecho do sinal da característica acústica RMS extraída da música “*The Flight of the Bumblebee*” de Nikolai Rimsky-Korsakov.



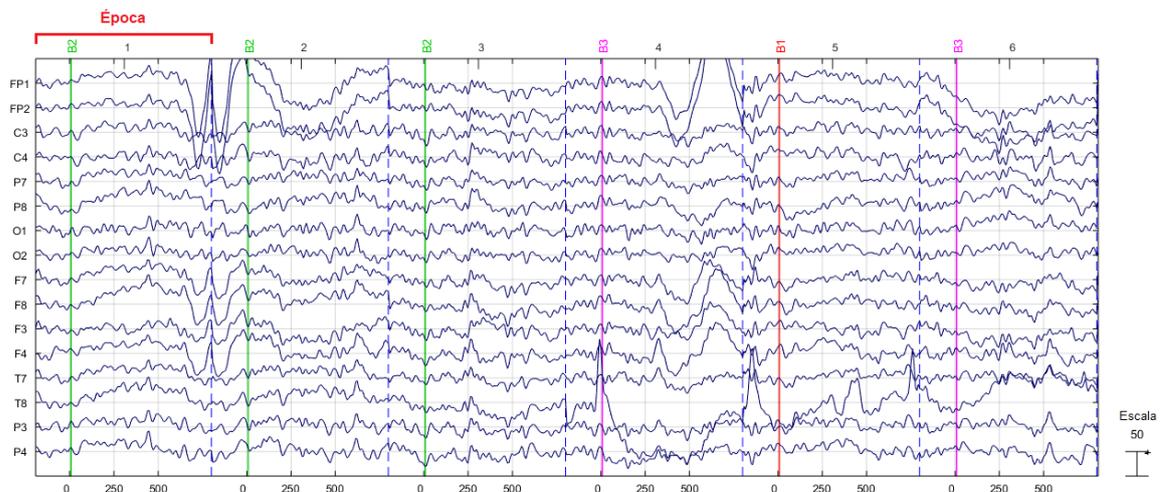
Fonte: Autor.

Este processo é importante pelo seguinte fato: Imagine que metaforicamente queiramos explorar uma mina de ouro. Neste caso, este material está presente em uma quantidade muito inferior em comparação aos demais tipos de rochas que podem estar presentes no local. Ou seja, o ouro está escondido no meio de tantos outros materiais. Da mesma forma, temos uma grande quantidade de dados referentes aos sinais de EEG e de áudio, enquanto a informação que queremos obter está escondida no meio de tantos dados. O *trigger*, neste caso, é a ferramenta que aponta para a região onde está o ouro. O ouro são os padrões neurais capazes de indicar se uma pessoa recebeu, ou não, treinamento musical em sua vida, após serem processados e analisados.

#### 4.2.4 Processamento dos sinais de EEG

Todos os sinais de EEG foram preprocessados usando um filtro rejeita-faixa de 60Hz, para remoção do sinal da rede elétrica, e um filtro passa-banda Butterworth de ordem 10 de 1-30 Hz, para remover o nível DC e deixar apenas as frequências de interesse para a análise do sinal de EEG. Este filtro foi gerado no ambiente Matlab, tendo como parâmetros: banda de passagem  $W_p = [1, 30]/fn$ , banda de rejeição  $W_s = [0.1, 50]/fn$  (onde  $fn$  corresponde à frequência de Nyquist), atenuação na banda de passagem  $A_p = 5$  dB e atenuação na banda de rejeição  $A_s = 50$  dB. Após a filtragem, é realizado o janelamento do sinal contínuo de EEG, separando-o em épocas demarcadas pelos *triggers*. A Figura 18 ilustra o janelamento de um sinal de EEG arbitrário demarcado por um conjunto de épocas (RIBEIRO, 2017).

Figura 18 – Ilustração do janelamento de um sinal de EEG arbitrário dividido em 6 épocas, delimitadas em 200 ms pré-estímulo e 800 ms pós-estímulo, para 3 tipos de estímulos distintos.



Fonte: Ribeiro, 2017, p.75.

Dois procedimentos simples e comuns aplicados aos sinais de EEG são o *averaging* e a *baseline correction* (LUCK, 2014). O *averaging* consiste simplesmente em somar um conjunto de épocas de EEG e então dividir pelo número total de épocas. Neste trabalho, as épocas iniciam 100 ms antes da ocorrência dos *triggers* e terminam 300 ms após o *trigger*. A *baseline correction* é alcançada subtraindo-se a tensão média do pré-estímulo do sinal dentro de cada época. Aqui, a *baseline correction* é definida como 100 ms antes da ocorrência dos *triggers*. Esse procedimento é necessário pois fatores como a hidratação da pele e cargas estáticas nos eletrodos podem causar efeitos de *offset* e *drift* nos sinais de EEG (LUCK, 2014).

Todos os sinais foram inspecionados visualmente e as épocas cujas amplitudes estavam acima de  $100 \mu\text{V}$  (em valores absolutos) foram removidas da análise. A toolbox EEGLab (versão 13.5.4b) (DELORME; MAKEIG, 2004) e a toolbox ERPLab (versão) (LOPEZ-CALDERON; LUCK, 2014) foram utilizadas para processar os dados de EEG no software Matlab R2015a.

#### 4.2.5 Análise estatística multivariada

Nesta quinta e última etapa, primeiramente é descrita a matriz de dados  $\mathbf{X}_r$  composta pelos valores das épocas médias do sinal de EEG em um instante  $t$  ms pós-estímulo (pré-definido) para cada característica acústica  $r$  selecionada, sendo  $r = 1, 2, \dots, sf$ , de tal forma que o sinal de entrada correspondente possa ser representado como um ponto no espaço multivariado de alta dimensão, como segue:

$$\mathbf{X}_r = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_N \end{bmatrix} = \begin{bmatrix} x_{11} & \dots & x_{1n} \\ x_{21} & \dots & x_{2n} \\ \vdots & \vdots & \vdots \\ x_{N1} & \dots & x_{Nn} \end{bmatrix}, \quad (2)$$

em que  $N$  é o número total de voluntários (ou sujeitos) e  $n$  é o número total de eletrodos ( $n = 16$ , aqui).

Como dito anteriormente, Poikonen et al. (2016) foram capazes de identificar os componentes N1 e P2 do ERP utilizando os *triggers* identificados na análise do áudio. Sabe-se que o componente N1 ocorre na vizinhança da região temporal de 100 ms pós-estímulo e o componente P2 próximo de 200 ms. É comum buscar o pico da onda em uma janela temporal (correspondente ao componente investigado) e utilizar esta informação para compor a análise. No entanto, esta abordagem faz com que haja variabilidade temporal entre os sujeitos. Além disso, a ocorrência dos componentes de ERP não é determinística. Por estas razões definiu-se aqui um instante de tempo pré-definido para se analisar os sinais de EEG comum para todos os voluntários, assumindo  $t = 100$  ms pós-estímulo correspondendo ao componente N1 do ERP.

De acordo com essa representação dos dados, é assumido que esta análise estatística multivariada dos sinais de EEG é capaz de compreender a complexidade da atividade cerebral em termos das  $N$  amostras e dos  $n$  eletrodos simultaneamente, podendo incluir tanto os

potenciais elétricos registrados em intervalos de tempo pré-definidos (como os realizados nos experimentos) quanto dados sintetizados ou adquiridos em um espaço multilinear.

A Análise de Componentes Principais (PCA) e a Análise Linear Discriminante, ou *Linear Discriminant Analysis* (LDA) (JOHNSON; WICHERN, 2007) são exploradas como dois métodos de estatística multivariada pra compreender como a informação se altera no espaço original dos dados de EEG, permitindo analisar não apenas as mudanças mais expressivas (maior variância), mas também as mais discriminantes (maior separabilidade), de acordo com as respostas neurais refletidas pelas características acústicas extraídas do áudio utilizado como estímulo para os voluntários.

#### 4.2.5.1 Análise de componentes principais (PCA)

A Análise de Componentes Principais (em inglês, *Principal Component Analysis - PCA*) foi utilizada para destacar as mudanças mais expressivas em termos da variância total da informação dos potenciais elétricos registrados.

O PCA calcula a decomposição espectral da matriz de correlação da matriz  $\mathbf{X}_r$  de tamanho  $N \times n$ , descrita na Equação 2. A matriz de transformação do PCA foi composta selecionando os  $M$  autovetores com autovalores não-nulos, tal que  $M \leq n$ . Os dados foram projetados nos  $[\mathbf{p}_{1,r}, \mathbf{p}_{2,r}, \dots, \mathbf{p}_{M,r}]$  autovetores rearranjados em ordem decrescente pelos seus autovalores correspondentes, ou seja,  $\lambda_{1,r} \geq \lambda_{2,r} \geq \dots \geq \lambda_{M,r}$ , de acordo com a direção de maior variância de  $\mathbf{X}_r$ .

Cada hiperplano  $\mathbf{p}_{m,r}$ , em que  $m = 1, 2, \dots, M$ , pode ser utilizado para descrever um mapa espacial das regiões cerebrais que mais variam nos dados, movendo de um lado do eixo da componente principal (ou dimensão) para o outro lado (COOTES; EDWARDS; TAYLOR, 1998; DAVATZIKOS, 2004). Portanto, é possível navegar ao longo de tais dimensões para capturar e compreender as mudanças mais expressivas na matriz de dados  $\mathbf{X}_r$ . Este método de navegação no hiperplano do PCA (COOTES; EDWARDS; TAYLOR, 1998; SATO et al., 2008; XAVIER et al., 2015; GREGORI; SANCHES; THOMAZ, 2017) pode ser matematicamente descrito como

$$\mathbf{y}_{m,h} = \bar{\mathbf{x}}_r + (h\sqrt{\lambda_{m,r}}) \times \mathbf{p}_{m,r}, \quad (3)$$

em que  $h \in \{-3, -2, -1, 0, 1, 2, 3\}$ ,  $m$  é o índice da componente principal correspondente para a navegação, e  $\bar{\mathbf{x}}_r$  é o vetor de média global  $n$ -dimensional de  $\mathbf{X}_r$  para cada característica acústica  $r$  selecionada.

#### 4.2.5.2 *Análise linear discriminante (LDA)*

A Análise Linear Discriminantes (em inglês, *Linear Discriminant Analysis - LDA*) foi utilizada para identificar as dimensões mais discriminantes para separar os dois grupos de interesse (ou classes), ou seja, músicos e não-músicos, maximizando a separabilidade inter-classes enquanto se minimiza a variabilidade intra-classe.

Sejam as matrizes inter-classe e intra-classe  $\mathbf{S}_b$  e  $\mathbf{S}_w$  definidas, respectivamente, como (FUKUNAGA, 1990)

$$\mathbf{S}_b = \sum_{i=1}^c N_i (\bar{\mathbf{x}}_i - \bar{\mathbf{x}}_r)^T (\bar{\mathbf{x}}_i - \bar{\mathbf{x}}_r), \quad (4)$$

$$\mathbf{S}_w = \sum_{i=1}^c \sum_{j=1}^{N_i} (\mathbf{x}_{i,j} - \bar{\mathbf{x}}_i)^T (\mathbf{x}_{i,j} - \bar{\mathbf{x}}_i), \quad (5)$$

em que  $\mathbf{x}_{i,j}$  é a amostra  $n$ -dimensional  $j$  da classe  $i$ ,  $N_i$  é o número de amostras de treinamento da classe  $i$  ( $N_1 = 13$  e  $N_2 = 13$ , aqui),  $c$  representa o número total de classes ( $c = 2$ , aqui) e  $\bar{\mathbf{x}}_i$  é o vetor de média das amostras de cada classe.

O principal objetivo do LDA é encontrar um hiperplano de projeção  $\mathbf{w}_r$  que maximize a razão do determinante de  $\mathbf{S}_b$  e  $\mathbf{S}_w$  (critério de Fisher), formulado como:

$$\mathbf{w}_r = \arg \max_{\mathbf{w}} \frac{|\mathbf{w}^T \mathbf{S}_b \mathbf{w}|}{|\mathbf{w}^T \mathbf{S}_w \mathbf{w}|}. \quad (6)$$

O critério de Fisher é maximizado quando o hiperplano  $\mathbf{w}_r$  é composto pelo autovetor principal de  $\mathbf{S}_w^{-1} \mathbf{S}_b$  com autovalor correspondente não-nulo (FUKUNAGA, 1990; JOHNSON; WICHERN, 2007).

Analogamente ao PCA, esse hiperplano  $\mathbf{w}_r$  pode ser utilizado para descrever um mapa espacial das regiões cerebrais mais discriminantes, movendo de um lado da dimensão de discriminância para o outro lado (DAVATZIKOS, 2004; SATO et al., 2008). Portanto, é possível navegar ao longo de tal dimensão para capturar e compreender as mudanças mais discriminantes na matriz de dados  $\mathbf{X}_r$ , assumindo que essa dispersão assuma uma distribuição Gaussiana

(SATO et al., 2008; XAVIER et al., 2015; GREGORI; SANCHES; THOMAZ, 2017). Esse método de navegação no hiperplano do LDA pode ser matematicamente descrito como

$$\mathbf{z}_{i,j} = \bar{\mathbf{x}}_r + (j\sigma_i + \bar{x}_i) \times \mathbf{w}_r, \quad (7)$$

em que  $j \in \{-1, 0, 1\}$ ,  $i$  é novamente o grupo de amostragem correspondente,  $\bar{\mathbf{x}}_r$  é o vetor de média global  $n$ -dimensional de  $\mathbf{X}_r$ , e  $\sigma_i$  e  $\bar{x}_i$  são, respectivamente, o desvio padrão e a média de cada grupo amostrado no espaço do LDA para cada característica acústica  $r$  selecionada.

#### 4.2.5.3 Classificação

Dado o número limitado de amostras disponível, adota-se os métodos *leave-one-out* e de ressubstituição para estimar respectivamente os limites inferiores (conjunto de teste) e superiores (conjunto de treinamento) das dimensões de maior discriminância, usando um classificador de distância Euclidiana com relação ao vetor de média de cada classe para decidir se os dados projetados se assemelham mais ao grupo de músicos ou não-músicos.

De acordo com Fukunaga (1990), o erro de classificação de uma estimativa baseada em poucas amostras, como as descritas neste trabalho, pode ser determinado em função de dois conjuntos de dados: o conjunto de treinamento ( $\Omega_d$ ) e o conjunto de teste ( $\Omega_t$ ). Fukunaga (1990) demonstra que os valores superiores e inferiores esperados ( $E\{\cdot\}$ ) do erro da classificação da população ( $\Omega$ ) podem ser calculados pelos seguintes limites

$$E\{\varepsilon(\Omega_d, \Omega_d)\} \leq \varepsilon(\Omega, \Omega) \leq E\{\varepsilon(\Omega_d, \Omega_t)\}. \quad (8)$$

Esta fórmula descreve que de um lado, o termo mais à direita  $\varepsilon(\Omega_d, \Omega_t)$ , é calculado ao se gerar dois conjuntos de amostras diferentes, usando  $\Omega_d$  para o treinamento e  $\Omega_t$  para teste. Este procedimento é chamado de método *holdout*, aqui implementado como o método *leave-one-out*, e seu valor esperado retorna o limite superior do erro de classificação. Do outro lado, o termo mais à esquerda  $\varepsilon(\Omega_d, \Omega_d)$  é obtido usando o mesmo conjunto de amostras para treinamento e teste e seu valor esperado retorna o limite inferior do erro. Este é o chamado método de ressubstituição.

Baseando-se neste conceito, é utilizado  $1 - \varepsilon(\cdot)$  para estimar os limites superiores e inferiores da acurácia de classificação obtidas.

## 4.3 RESULTADOS

### 4.3.1 Participantes

Dois grupos de 13 sujeitos fizeram parte do experimento. Os músicos (7 homens e 6 mulheres) eram todos amadores, com idade entre 18 e 45 anos (média 26.8 anos; todos destros). Eles iniciaram suas atividades musicais entre 4 e 17 anos (média 17.1 anos) e atualmente praticam seus instrumentos 4 horas por semana em média, com estilos musicais distintos (clássico, pop e rock). Todos os voluntários músicos receberam mais de 2 anos de treinamento formal em música, mas nenhum possui formação acadêmica nessa área. Os não-músicos (10 homens e 3 mulheres) possuem idade entre 25 e 45 anos (média 31.2 anos; 9 destros). Dois dos não-músicos receberam menos do que um ano de treinamento formal em música (nenhum continua praticando) e os outros nunca receberam treinamento formal em música. A Tabela 1 apresenta uma descrição mais detalhada dos voluntários que participaram dos experimentos. Todos os participantes assinaram o termo de consentimento livre e esclarecido para realização dos experimentos.

### 4.3.2 Estímulo

Como estímulo foi utilizada a música clássica “Dança Húngara No.5” de Johannes Brahms, apresentada via fones intra-auriculares. A performance foi executada pela Orquestra Sinfônica Fulda, conduzida por Simon Schindler (2000) no *Fürstensaal des Stadtschlosses*<sup>2</sup>. O final do áudio, correspondendo aos aplausos, foi substituído por silêncio utilizando o software Audacity (versão 2.2.2), resultando em um sinal de áudio de 3.2 minutos de duração. Esta música foi escolhida por estar no contexto de música clássica abordada na literatura afim (ALLURI et al., 2012, 2013; POIKONEN et al., 2016), por possuir muitas variações de andamento e dinâmica, gerando momentos contrastantes no áudio, e por ser uma música muito conhecida. Todos os voluntários declararam que conheciam a música após ouvi-la. Durante o experimento, todos os voluntários foram instruídos a ouvirem atentamente a música em posição mais imóvel possível, enquanto os sinais de EEG eram registrados. Não foi necessário realizar nenhum tipo de tarefa além da escuta musical.

<sup>2</sup>Disponível em: [https://imslp.org/wiki/21\\_Hungarian\\_Dances\\_\(Orchestra\)%2C\\_WoO\\_1\\_\(Brahms%2C\\_Johannes\)](https://imslp.org/wiki/21_Hungarian_Dances_(Orchestra)%2C_WoO_1_(Brahms%2C_Johannes))

Tabela 1 – Descrição dos voluntários que participaram dos experimentos.

ID	Sexo	Idade	Destro/Canhoto	Instrumento	Anos de prática	Horas de prática semanal
NM1	Masculino	27 anos	Destro	-	-	-
NM2	Masculino	46 anos	Canhoto	-	-	-
NM3	Masculino	40 anos	Canhoto	-	-	-
NM4	Masculino	30 anos	Destro	-	-	-
NM5	Masculino	25 anos	Destro	-	-	-
NM6	Feminino	25 anos	Destra	-	-	-
NM7	Masculino	42 anos	Canhoto	-	-	-
NM8	Masculino	24 anos	Destro	-	-	-
NM9	Masculino	28 anos	Destro	-	-	-
NM10	Masculino	25 anos	Canhoto	-	-	-
NM11	Masculino	25 anos	Destro	-	-	-
NM12	Feminino	35 anos	Destra	-	-	-
NM13	Feminino	34 anos	Destra	-	-	-
M1	Masculino	26 anos	Destro	Bateria	10 anos	-
M2	Masculino	18 anos	Destro	Flauta transversal	7 anos	4hrs
M3	Feminino	23 anos	Destro	Contrabaixo acústico	13 anos	4hrs
M4	Feminino	45 anos	Destro	Saxofone	28 anos	4hrs
M5	Feminino	25 anos	Destro	Violino	14 anos	4hrs
M6	Masculino	26 anos	Destro	Bateria	20 anos	8hrs
M7	Masculino	25 anos	Destro	Piano	21 anos	35hrs
M8	Masculino	31 anos	Destro	Violão	17 anos	-
M9	Masculino	26 anos	Destro	Guitarra	10 anos	-
M10	Feminino	22 anos	Destro	Violão	7 anos	8hrs
M11	Feminino	28 anos	Destro	Canto/Violino	13 anos	-
M12	Feminino	28 anos	Destro	Violino	18 anos	4hrs
M13	Masculino	26 anos	Destro	Saxofone	13 anos	4hrs

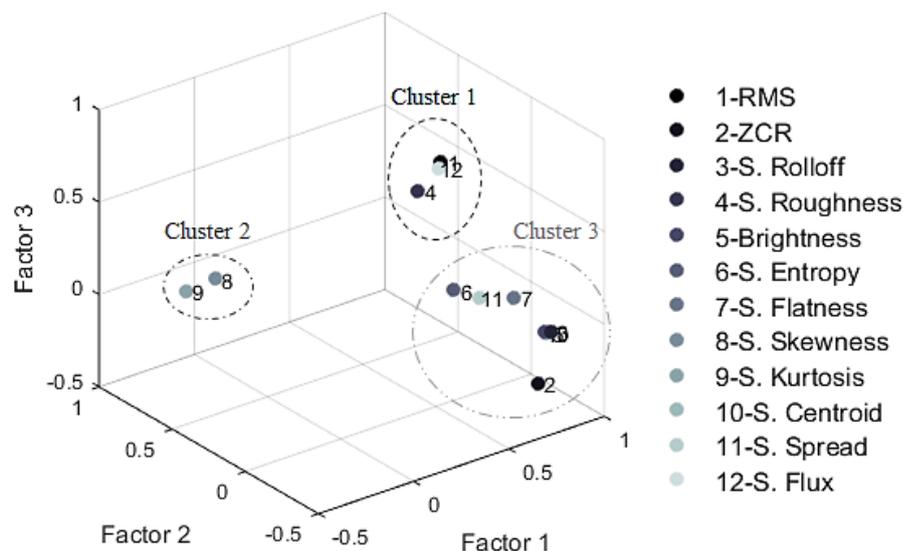
Fonte: Ribeiro e Thomaz, 2017.

### 4.3.3 Análise do sinal de áudio

Os valores das cargas fatoriais resultantes do FA são apresentados na Figura 19 e na Tabela 2. É possível observar claramente que as características acústicas extraídas do áudio formam 3 agrupamentos (*clusters*), revelando a relação entre essas características. A formação destes agrupamentos indica que existem similaridades entre a informação acústica expressada pelas características acústicas extraídas. Desta forma, foram selecionadas as características que possuíam a maior carga fatorial em cada fator, sendo essas: RMS (*cluster* 1), Spectral Kurtosis

(*cluster 2*) e Spectral Rolloff (*cluster 3*). Note que para o Fator 2, o Spectral Skewness possuía maior carga fatorial, porém, como o número de *triggers* encontrados para essa característica foi apenas 3, foi decidido utilizar o Spectral Kurtosis no seu lugar.

Figura 19 – Carga fatorial das características acústicas extraídas do sinal de áudio, sendo possível observar a formação de *clusters* entre as seguintes características: Cluster 1 — 1 (RMS), 4 (Spectral Roughness) e 12 (Spectral Flux); Cluster 2 — 8 (Spectral Skewness) e 9 (Spectral Kurtosis); Cluster 3 — 2 (ZCR), 3 (Spectral Rolloff), 5 (Brightness), 6 (Spectral Entropy), 7 (Spectral Flatness), 10 (Spectral Centroid) e 11 (Spectral Spread).



Fonte: Ribeiro e Thomaz, 2019.

Para as características acústicas selecionadas foram encontrados 9 *triggers* para o RMS, 7 para o Spectral Rolloff e 8 para o Spectral Kurtosis ao longo da música utilizada como estímulo. Como é possível observar na Figura 20, existem similaridades entre os espectrogramas pertencentes ao mesmo *cluster*. Os espectrogramas descrevem melhor a ocorrência dos *triggers* em todas as características acústicas do que os sinais de áudio no domínio do tempo, pois apresentam visualmente as distinções entre regiões com baixa e alta energia que estão relacionadas ao contraste acústico proposto por Poikonen et al. (2016). Além disso, embora existam *triggers* que ocorram em instantes diferentes dentro do mesmo *cluster*, a grande maioria ocorre simultaneamente. Esses resultados indicam que as características extraídas são redundantes entre si, especialmente dentro do mesmo *cluster*. Neste caso foram utilizados os mesmos parâmetros

Tabela 2 – Carga fatorial das características acústicas extraídas do sinal de áudio.

Característica	F1	F2	F3
RMS	0.6291	0.1337	<b>0.7203</b>
ZCR	0.7851	-0.3235	-0.3240
S. Rolloff	<b>0.9584</b>	-0.1939	-0.1629
S. Roughness	0.3887	-0.0242	0.7185
Brightness	0.9471	-0.1643	-0.1790
S. Entropy	0.8779	0.3712	-0.1710
S. Flatness	0.9329	0.0290	-0.0764
S. Skewness	0.0995	0.9847	-0.1210
S. Kurtosis	-0.0761	<b>0.9617</b>	-0.1139
S. Centroid	0.9563	-0.1484	-0.1975
S. Spread	0.9247	0.2493	-0.1738
S. Flux	0.5880	0.0931	0.7174

Fonte: Ribeiro e Thomaz, 2019.

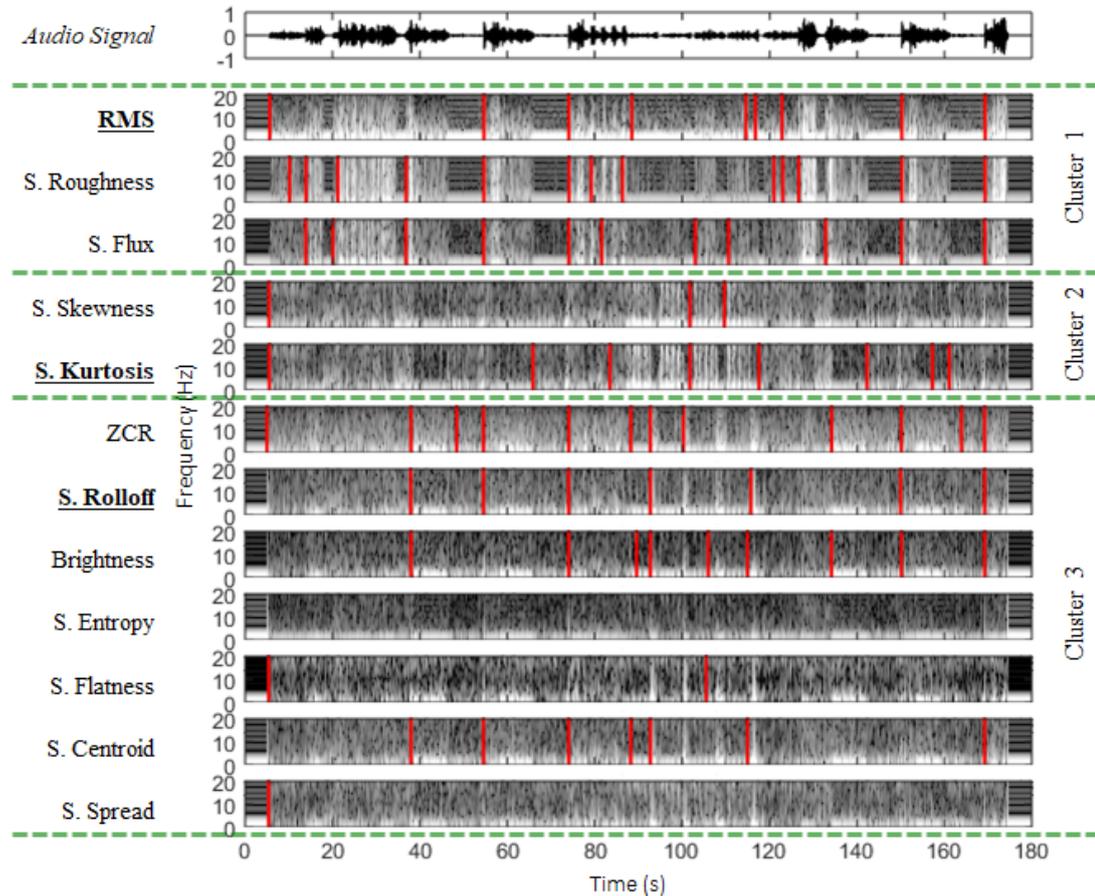
para a seleção dos *triggers*, no entanto, é possível alterá-los para obter resultados diferentes, considerando o compromisso que deve ser feito com relação a quantidade de *triggers* e o contraste acústico gerado.

#### 4.3.4 Análise do sinal de EEG

Os mapas topográficos foram gerados pelo método de navegação no hiperplano nas quatro dimensões mais expressivas do PCA, de acordo com a Equação (3), e na dimensão discriminante do LDA, de acordo com a Equação (7), para cada característica acústica selecionada. A navegação nestes hiperplanos revela visualmente as mudanças no potencial elétrico que foram registradas dos voluntários pelo equipamento de EEG durante a tarefa de escuta musical, assim como a distribuição do potencial elétrico ao longo das dimensões correspondentes neste espaço multivariado. Desta forma, conseguimos identificar visualmente as diferenças nos padrões de ativação que caracterizam os grupos estudados, bem como as regiões de transições que separam cognitivamente as amostras de músicos e não-músicos referentes ao processamento das características acústicas selecionadas.

Os mapas topográficos e as projeções dos dados gerados pelo PCA da característica acústica Spectral Rolloff são apresentadas na Figura 21, sendo possível observar mudanças nas respostas cerebrais ao longo de cada componente principal (PC), revelando diferenças nos potenciais elétricos em diversas partes do córtex, especialmente na área frontal. Entretanto, quando os dados são projetados nos eixos de cada PC separadamente para classificação, não

Figura 20 – Espectrogramas de cada característica acústica extraída do sinal de áudio separadas em clusters (linhas verdes tracejadas), apresentando os *triggers* selecionados (linhas vermelhas) e a característica acústica selecionada (em negrito e sublinhado) para cada cluster.

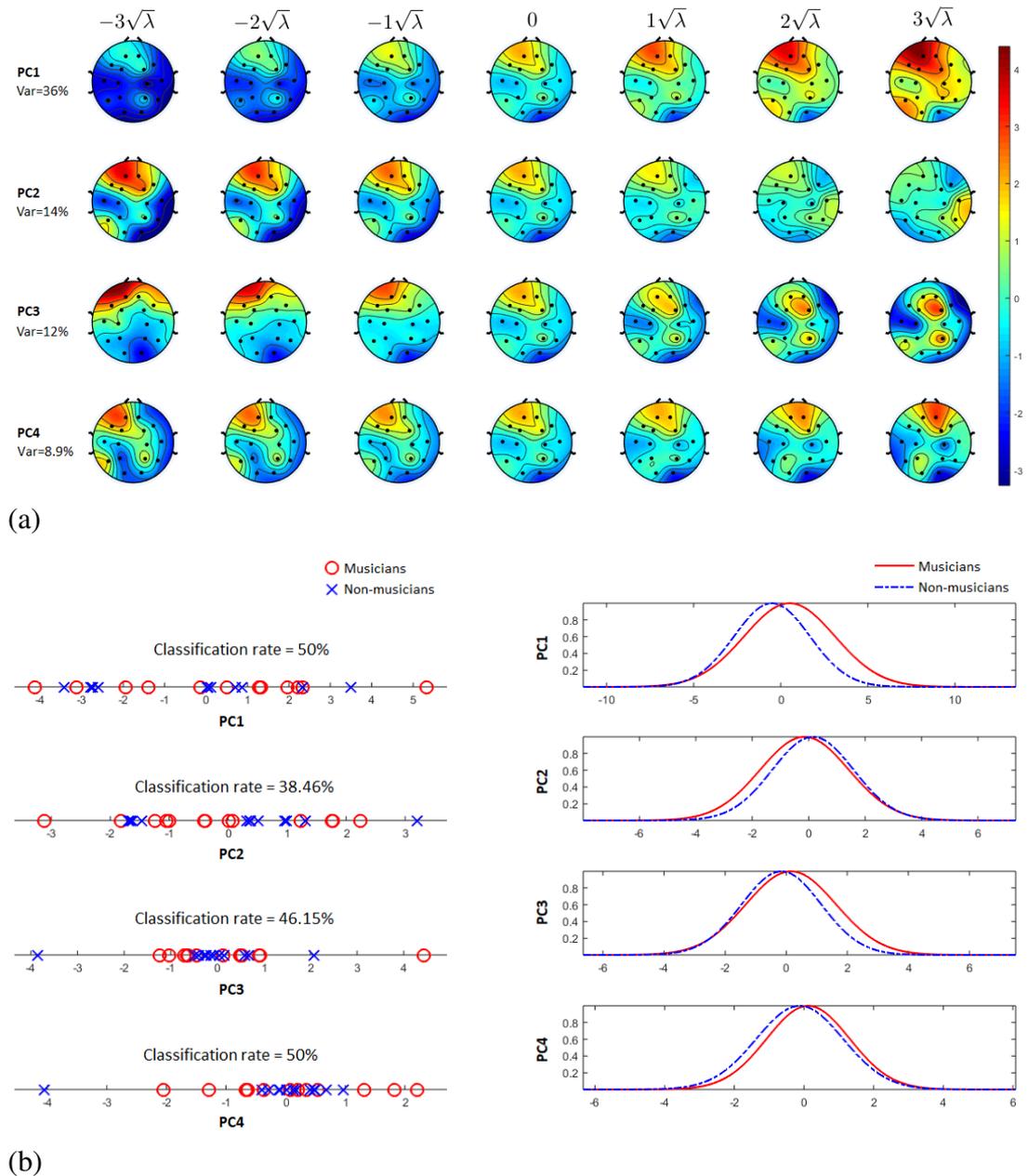


Fonte: Ribeiro e Thomaz, 2019.

é possível observar uma distinção entre os grupos de músicos e não-músicos. Diferente do PCA, a Figura 22 apresenta a navegação no eixo discriminante do LDA, mostrando diferenças entre os grupos nas áreas frontais, sendo que os não-músicos apresentam um potencial elétrico positivo distribuído na área frontal, enquanto os músicos apresentam maior atividade na área pré-frontal. Tais diferenças não podem ser consideradas discriminantes, como pode ser visto pela projeção e distribuição dos dados, pois apresentam uma taxa de classificação que varia entre 15.4% (conjunto de teste) e 74.5% (conjunto de treinamento).

Adicionalmente, a Figura 23 apresenta os mapas topográficos e a projeção dos dados gerados pelo PCA da característica acústica RMS, sendo possível observar que as mudanças acontecem principalmente nas áreas frontais do córtex, mas novamente não é possível ver uma distinção entre os grupos de músicos e não-músicos nos dados projetados em cada PC. Seme-

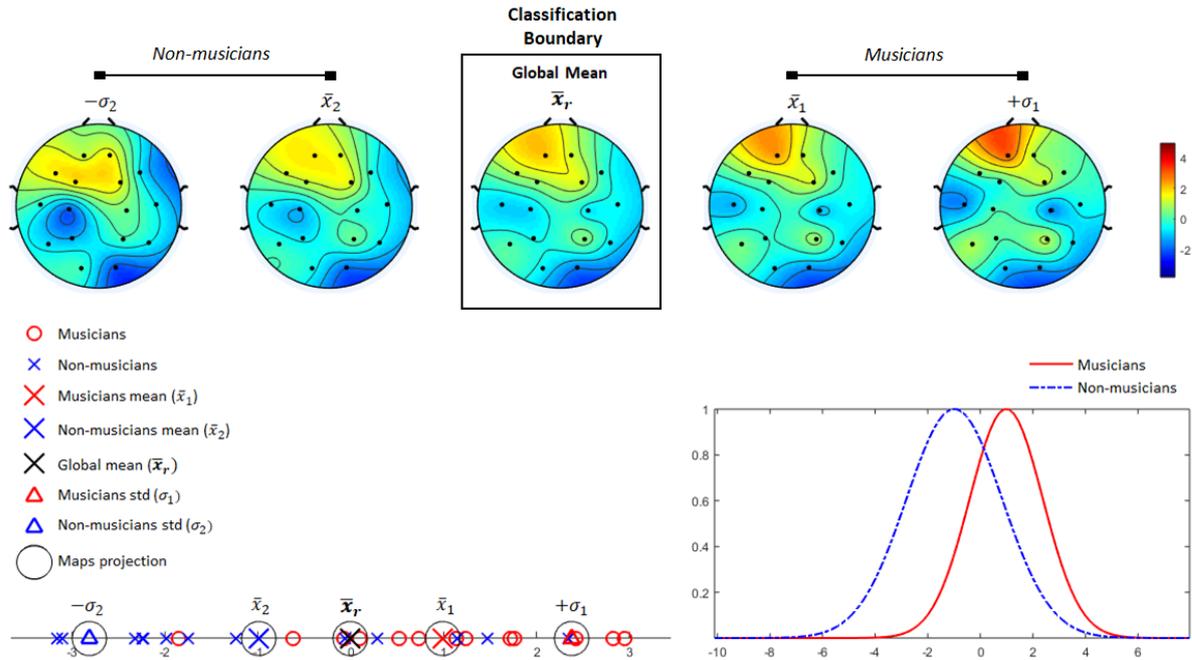
Figura 21 – Navegação no hiperplano das quatro primeiras componentes principais (PCs), ordenadas da maior para menor variância, da característica acústica **Spectral Rolloff** gerados pelo PCA no instante 100 ms pós-estímulo: (a) Mapas topográficos de cada PC; (b) Projeção dos dados em cada PC.



Fonte: Ribeiro e Thomaz, 2019.

lhantemente, a Figura 24 apresenta a navegação no eixo discriminante da característica RMS, apresentando visualmente uma diferença sutil no potencial elétrico entre os grupos, porém estas diferenças são estatisticamente mais discriminantes do que a característica anterior (Spectral

Figura 22 – Navegação no hiperplano discriminante da característica acústica **Spectral Rolloff** gerada pelo LDA no instante 100 ms pós-estímulo: (Cima) Reconstrução do mapa topográfico médio durante a navegação no eixo discriminante entre o grupo de músicos (direita) e não-músicos (esquerda); (Baixo) Projeção dos dados no eixo discriminante. *Acurácia: 15.4% (conjunto de teste) e 74.5% (conjunto de treinamento).*



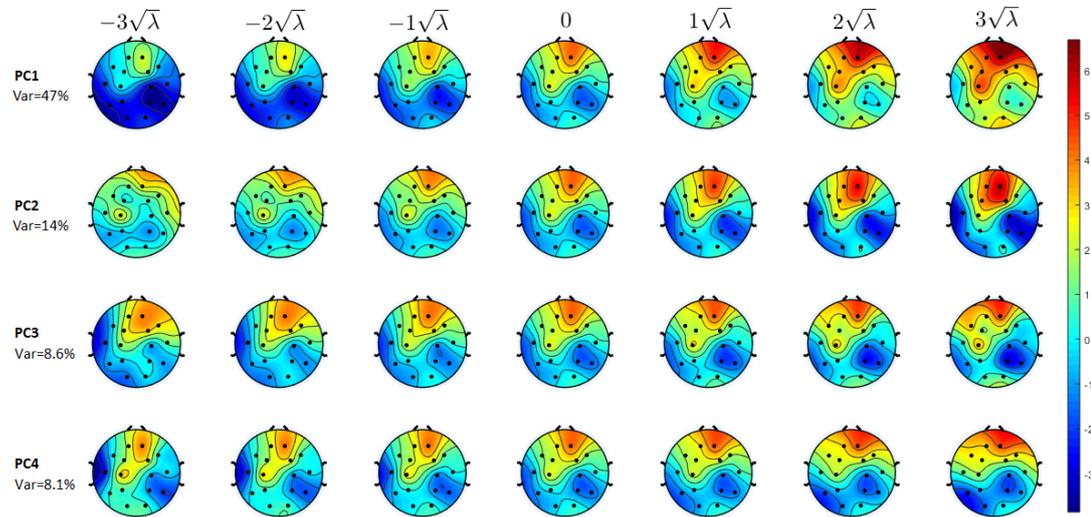
Fonte: Ribeiro e Thomaz, 2019.

Rolloff), com taxa de classificação entre 53.8% (conjunto de teste) a 88.0% (conjunto de treinamento).

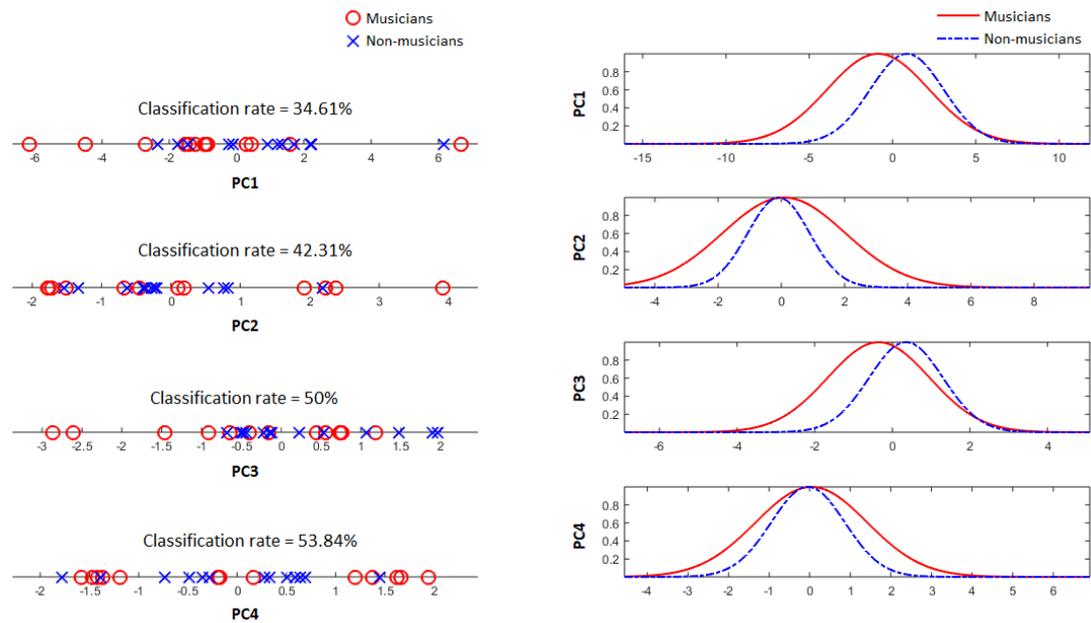
Embora, analogamente aos resultados anteriores, o resultado do PCA para a característica Spectral Kurtosis, apresentada na Figura 25, não apresente distinção entre os grupos, essa característica apresenta, na Figura 26, uma clara distinção no potencial elétrico entre os grupos de interesse. Nesse caso, os não-músicos apresentam alta atividade na região pré-frontal enquanto os músicos apresentam maior atividade na região temporal e um potencial elétrico negativo distribuído na área parietal. A distribuição dos dados exibe claramente uma separação entre os grupos, com taxa de classificação entre 69.2% (conjunto de teste) e 93.8% (conjunto de treinamento).

Para comparação, foi implementada a mesma abordagem para discriminar os grupos, porém utilizando um número reduzido de eletrodos em vez de utilizar todos eles, seguindo o seguinte arranjo: 6 eletrodos (F3, F4, C3, C4, P3, P4); 4 eletrodos (F3, F4, C3, C4); 2 eletrodos (F3, F4) e 2 eletrodos (C3, C4). A Tabela 3 apresenta a taxa de classificação do LDA para

Figura 23 – Navegação no hiperplano das quatro primeiras componentes principais (PCs), ordenadas da maior para menor variância, da característica acústica **Root Mean Square (RMS)** gerados pelo PCA no instante 100 ms pós-estímulo: (a) Mapas topográficos de cada PC; (b) Projeção dos dados em cada PC.



(a)

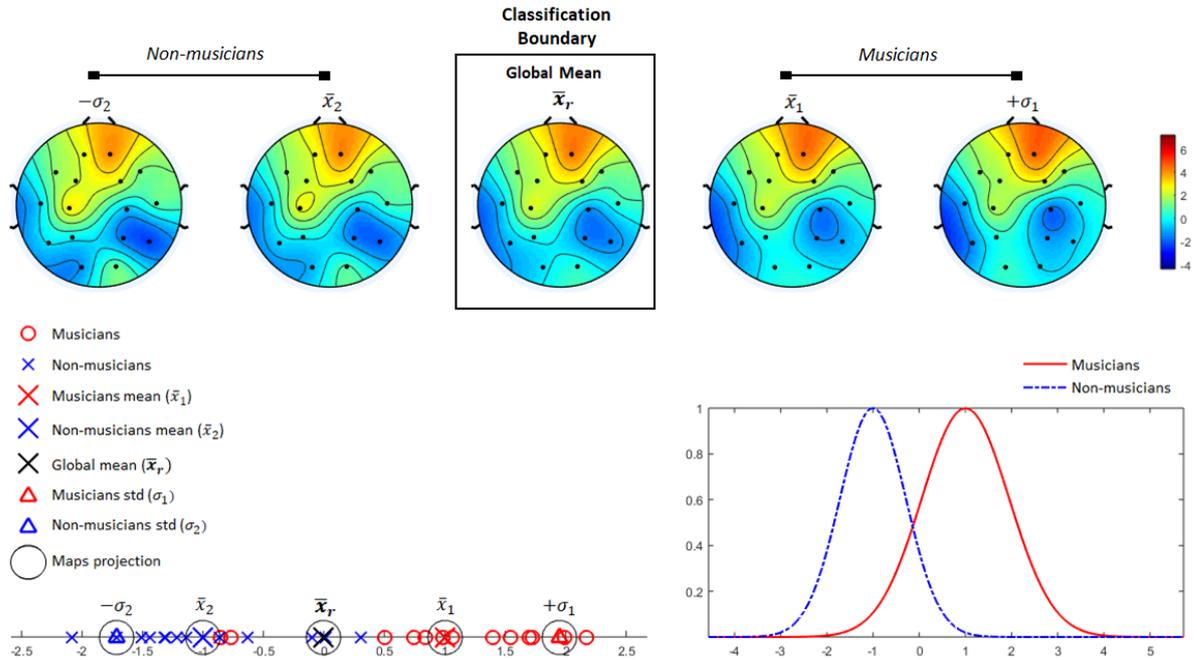


(b)

Fonte: Ribeiro e Thomaz, 2019.

cada característica acústica e a Tabela 4 a taxa de classificação detalhada utilizando todos os eletrodos.

Figura 24 – Navegação no hiperplano discriminante da característica acústica **Root Mean Square (RMS)** gerada pelo LDA no instante 100 ms pós-estímulo: (Cima) Reconstrução do mapa topográfico médio durante a navegação no eixo discriminante entre o grupo de músicos (direita) e não-músicos (esquerda); (Baixo) Projeção dos dados no eixo discriminante. *Acurácia: 53.8% (conjunto de teste) e 88.0% (conjunto de treinamento).*



Fonte: Ribeiro e Thomaz, 2019.

Tabela 3 – Taxa de classificação do LDA para todos os eletrodos, 6 eletrodos, 4 eletrodos e 2 eletrodos.

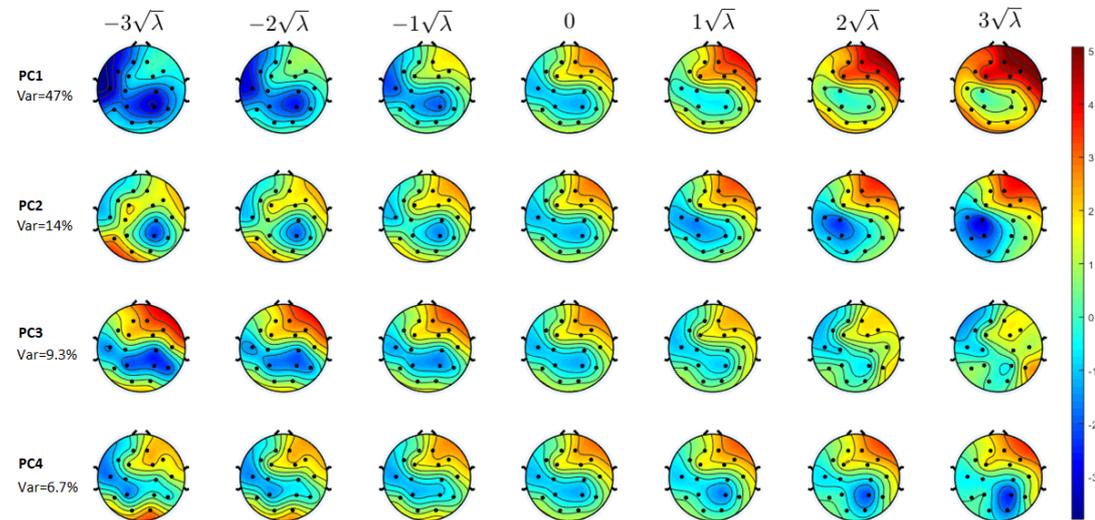
Características	Todos Eletrodos	6 Eletrodos	4 Eletrodos	2 Eletrodos (F3-F4)	2 Eletrodos (C3-C4)
<i>Teste</i>					
<i>Kurtosis</i>	69.2%	53.8%	65.4%	53.8%	73.1%
<i>Rms</i>	53.8%	57.7%	53.8%	61.5%	61.5%
<i>Rolloff</i>	15.4%	42.3%	50.0%	65.4%	53.8%
<i>Treino</i>					
<i>Kurtosis</i>	93.8%	72.9%	71.8%	54.0%	73.1%
<i>Rms</i>	88.0%	72.2%	65.4%	68.2%	67.5%
<i>Rolloff</i>	74.5%	68.5%	70.9%	69.4%	54.6%

Fonte: Ribeiro e Thomaz, 2019.

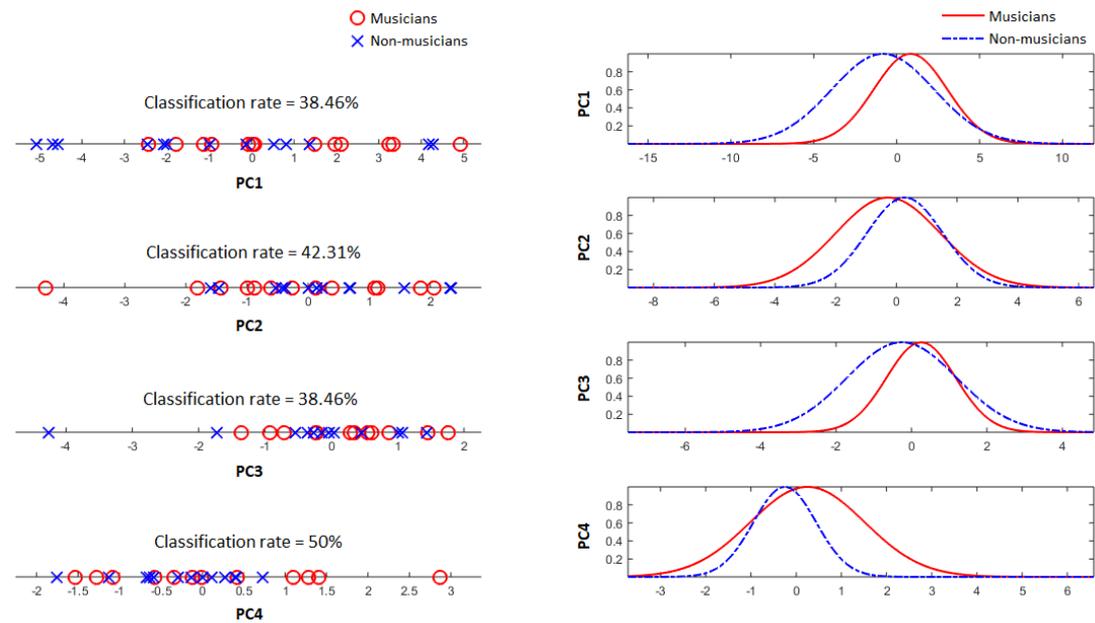
#### 4.4 DISCUSSÃO

Neste capítulo foi descrita uma metodologia de estatística multivariada para descrever as diferenças entre sujeitos categorizados como músicos e não-músicos. Os principais resultados

Figura 25 – Navegação no hiperplano das quatro primeiras componentes principais (PCs), ordenadas da maior para menor variância, da característica acústica **Spectral Kurtosis** gerados pelo PCA no instante 100 ms pós-estímulo: (a) Mapas topográficos de cada PC; (b) Projeção dos dados em cada PC.



(a)

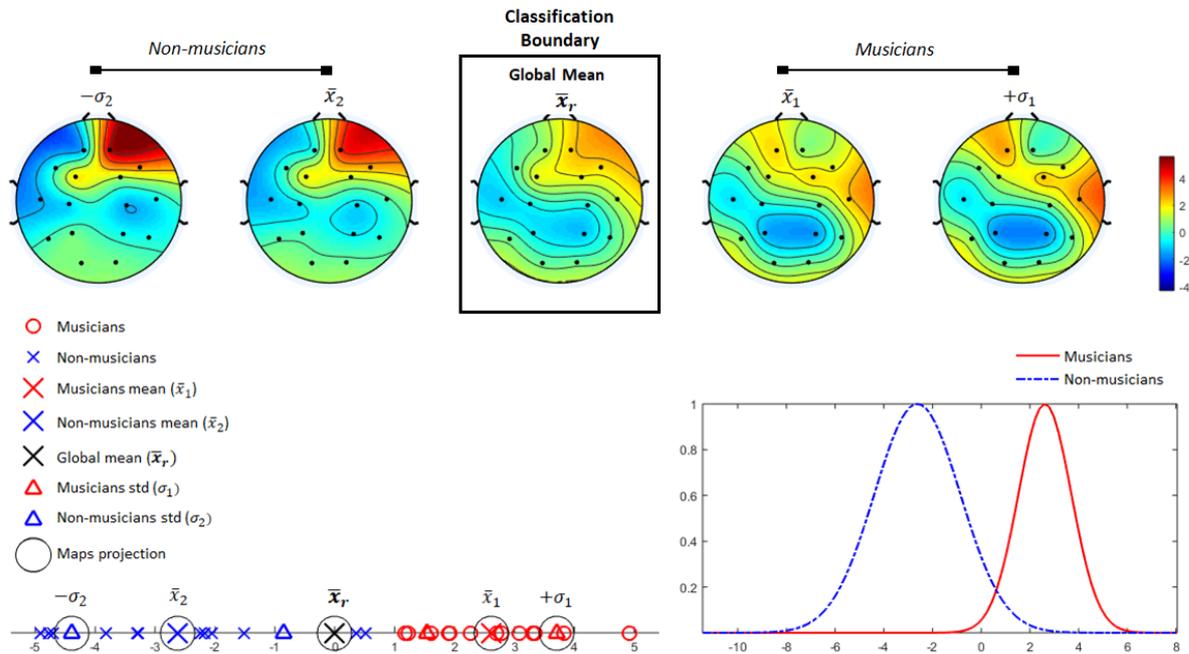


(b)

Fonte: Ribeiro e Thomaz, 2019.

podem ser assim resumidos: (1) a fim de investigar os padrões de ativação neural evocados pela escuta musical nos instantes demarcados pelas características acústicas, foi demonstrado que nem todas as 12 características extraídas do áudio são necessárias para a análise, devido à redundância encontrada pelo FA, que agrupa todas essas 12 características em 3 *clusters*

Figura 26 – Navegação no hiperplano discriminante da característica acústica **Spectral Kurtosis** gerada pelo LDA no instante 100 ms pós-estímulo: (Cima) Reconstrução do mapa topográfico médio durante a navegação no eixo discriminante entre o grupo de músicos (direita) e não-músicos (esquerda); (Baixo) Projeção dos dados no eixo discriminante. *Acurácia: 69.2% (conjunto de teste) e 93.8% (conjunto de treinamento).*



Fonte: Ribeiro e Thomaz, 2019.

Tabela 4 – Taxa de classificação detalhada do LDA utilizando todos os eletrodos.

Características	Acurácia	Sensibilidade	Especificidade	distribuição F
<i>Teste</i>				
<i>Kurtosis</i>	69.2	69.2	69.2	69.2
<i>Rms</i>	53.8	61.5	46.2	57.1
<i>Rolloff</i>	15.4	30.8	0	26.7
<i>Treino</i>				
<i>Kurtosis</i>	93.8	87.7	100	93.4
<i>Rms</i>	88.0	89.8	86.2	88.2
<i>Rolloff</i>	74.5	71.4	77.5	73.7

Fonte: Ribeiro e Thomaz, 2019.

estatisticamente não-redundantes; (2) os padrões discriminantes requeridos para classificar os grupos de músicos e não-músicos existem predominantemente nas áreas frontais do córtex, mas existem diferenças entre as respostas neurais para as características acústicas que representam cada *cluster*, indicando que alguns aspectos da música são melhores para prever treinamento musical do que outros, como a Intensidade (representada pela característica acústica RMS) e

a Tonalidade (representada pela característica acústica S. Rolloff). No entanto, a informação que melhor predisse treinamento musical apresentava um aspecto implicitamente estatístico (representada pela característica acústica S. Kurtosis).

Estes resultados demonstram que o método proposto por Poikonen et al. (2016) para identificar *triggers* pode ser muito útil para diferenciar músicos e não-músicos, levando em conta o compromisso entre o número de *triggers* e o quão alto o contraste acústico pode ser definido. Considerando a similaridade entre as características acústicas nas cargas fatoriais (Figura 19), é interessante observar que cada característica acústica extraída corresponde a um dos *clusters* encontrado e os espectrogramas (Figura 20) claramente apresentam essas similaridades. A maioria dos *triggers* selecionados ocorre no mesmo instante dentro de cada *cluster*, embora algumas características apresentem mais *triggers* do que outras (RIBEIRO; THOMAZ, 2019). Esses *triggers* extras estão relacionados com os parâmetros metodológicos utilizados, já que foi decidido usar os mesmos parâmetros para todas as características neste trabalho. Uma abordagem alternativa poderia ser, em vez de selecionar apenas uma característica acústica para representar um determinado *cluster*, concatenar os *triggers* encontrados para todas as características dentro de um mesmo *cluster*. É importante destacar, no entanto, que esses resultados se referem à música clássica utilizada neste trabalho. Espera-se que esse comportamento de *cluster* seja similar entre músicas de mesmo gênero (RIBEIRO; THOMAZ, 2019) e o próximo capítulo apresenta uma análise profunda sobre essa questão.

Sabe-se que o treinamento musical promove diferenças entre músicos e não-músicos não apenas durante a escuta musical, mas também durante o estado de repouso (KLEIN et al., 2016). Há também evidências de diferenças entre músicos com diferentes tipos de estratégias de treinamento, gênero musical e experiências auditivas (VUUST et al., 2012; TERVANIEMI et al., 2016). Portanto, há uma grande variabilidade entre indivíduos para esse problema de classificação de dois grupos, mas mesmo com essas diferenças, ainda foi possível encontrar uma boa acurácia de classificação. Além disso, os resultados experimentais demonstraram que o método de navegação no hiperplano discriminante descrito e implementado aqui pode trazer maior compreensão sobre os padrões de ativação e as fronteiras de decisão dos grupos amostrados. Os mapas topográficos gerados por este método de navegação claramente apresentam padrões distintos de ativação neural entre músicos e não-músicos, aumentando o entendimento sobre os estados de transição entre os grupos analisados, apesar do número limitado de amostras disponível.

Os resultados da característica acústica Spectral Kurtosis (Figura 26) revelam uma classificação entre músicos e não-músicos superior do que as outras características analisadas, apresentando a melhor separação linear entre os grupos amostrados. É interessante observar que o Spectral Kurtosis e o Spectral Skewness foram agrupados no mesmo *cluster*, já que caracterizam a distribuição do espectro dos dados em termos da sua forma, juntamente com o Spectral Flatness. Ambos kurtosis e skewness provêm informação sobre o tipo e magnitude do desvio da normalidade (DECARLO, 1997) e têm sido demonstrado que uma mudança súbita no Spectral Kurtosis pode indicar transientes em sinais não-estacionários (DWYER, 1984; ANTONI, 2006). No entanto, embora essa característica acústica pareça adequada para discriminar músicos e não-músicos, sua relação com o que é de fato percebido pelos sujeitos permanece em aberto, sendo necessária uma investigação posterior.

A navegação nos hiperplanos das PCs mostra que embora o PCA encontre as direções que os dados de EEG mais variam, essas direções não são necessariamente aquelas que melhor discriminam os grupos amostrados. Enquanto isso, o LDA apresenta informações interessantes sobre os aspectos discriminantes da atividade neural de cada grupo. Ainda assim, esta navegação promove uma melhor descrição dos padrões de ativações neurais, das regiões de transição e das fronteiras de decisão que separam cognitivamente músicos de não-músicos. É possível antecipar que o uso de músicos profissionais para esse arcabouço possa promover melhor discriminação entre os grupos de interesse e, conseqüentemente, melhor desempenho na classificação. No entanto, uma limitação adicional desse experimento é que os sujeitos que participaram das gravações não foram equiparados de acordo com algumas variáveis, especialmente os músicos (e.g. gênero, início do treinamento musical, anos de prática, etc.).

Adicionalmente, ao se levar em consideração a atividade cerebral como um todo (considerando a informação registrada simultaneamente em todos os eletrodos de EEG) durante a tarefa auditiva, pretende-se reforçar a importância de avaliar as possíveis interações entre diferentes regiões cerebrais, porque tais inter-relações podem ser úteis para separar os grupos investigados (DAVATZIKOS, 2004; FRISTON; ASHBURNER, 2004; THOMAZ et al., 2007). A abordagem de estatística multivariada descrita permite ter uma nova perspectiva para identificar novas regiões discriminantes potenciais para caracterizar musicalidade além das regiões específicas comumente abordadas. As diferenças estatísticas entre músicos e não-músicos registradas por essa abordagem revelam que as áreas cerebrais que melhor descrevem musicalidade existem predominantemente nas áreas frontais do córtex, corroborando com estudos anteriores (SAARI et al., 2018) sobre como o cérebro processa certas características acústicas da música, especial-

mente relacionadas com áreas motoras e auditivas. É interessante notar que a literatura fornece evidências de um processo de adaptação ao declínio cognitivo que resulta em um aumento da ativação cortical nas regiões frontais do cérebro (HO et al., 2012; DUSHANOVA; CHRISTOV, 2014; O'BRIEN; NIKJEH; LISTER, 2015).

Além disso, no trabalho de Poikonen et al. (2016) a atividade cerebral foi definida pela busca de um pico no sinal de EEG na vizinhança de 100 ms, correspondente a presença do componente N100. Neste experimento, apenas os músicos apresentaram esse fenômeno, impedindo de usar essa informação estratégica para compor a matriz de dados. Portanto, foi definido aqui o instante de 100 ms como um *time-stamp* igual e comparável para todos os sujeitos, músicos e não-músicos. No entanto, mesmo que fosse possível garantir que essa componente N100 acontecesse para todos os sujeitos, ainda estaríamos utilizando uma informação pré-definida baseada na resposta encontrada em regiões de interesse específicas (relacionadas aos eletrodos onde é esperado que este componente N100 aconteça), sem considerar a informação contida em outras regiões do cérebro. Acredita-se que essa limitação pode ser abordada utilizando algoritmos que sintetizem toda a informação da época média de cada sujeito (ROCHA et al., 2014), ou por meio de uma transformação de dados multilinear (LU; PLATANIOTIS; VENETSANOPOULOS, 2013).

A análise estatística multivariada implementada aqui foi realizada com um número limitado de amostras, dada a dimensionalidade original da matriz de dados e a dificuldade de recrutar voluntários, especialmente músicos profissionais, o que causa um impacto nas estimativas e na classificação das dimensões discriminantes. Isso se torna claro quando são utilizados menos eletrodos (Tabela 3) para comparar os resultados, sendo possível observar que as acurácias dos conjuntos de treinamento e teste são próximas, porém, na estimativa multivariada (ou seja, usando todos os eletrodos) verifica-se maior sensibilidade para tais escolhas, tendo em vista o maior intervalo entre os limiares inferiores (conjunto de teste) e superiores (conjunto de treinamento). Essa questão do número limitado de amostras pode ser abordado usando versões regularizadas do LDA, como o MLDA (THOMAZ et al., 2007; SATO et al., 2008), ou outros classificadores como o SVM (DAVATZIKOS, 2004). Outra possibilidade interessante seria abordar essa questão por meio de uma perspectiva de aprendizado em um subespaço multilinear (LU; PLATANIOTIS; VENETSANOPOULOS, 2013), considerando tempo, eletrodos e sujeitos como um tensor de terceira ordem, como mencionado no parágrafo anterior, podendo solucionar inclusive a limitação atual do *time-stamp* desse experimento.

## 5 PADRÕES SONOROS EM ÁUDIOS MUSICAIS

Existem muitas características que podem ser extraídas dos áudios, servindo para diferentes propósitos como a classificação do gênero das músicas, similaridade musical, detecção de emoções, entre outros, na área da Recuperação de Informação Musical (MIR). De acordo com Mitrović, Zeppelzauer e Breiteneder (2010), o desenvolvimento de sistemas de recuperação de áudio requer uma melhor representação desses sinais, o que significa que é necessário ter uma identificação mais apropriada das características acústicas utilizadas.

Como apresentado no capítulo anterior, uma aplicação diferente para estas características acústicas é utilizá-las como métrica para se analisar os sinais cerebrais de voluntários, seja por meio de equipamentos de fMRI ou de EEG. Trabalhos recentes apontam que é possível encontrar um comportamento de *cluster* entre características de baixo nível extraídas do sinal de áudio e tem-se a expectativa que esse comportamento seja similar entre músicas de mesmo gênero (FERREIRA; RIBEIRO; THOMAZ, 2019; RIBEIRO; THOMAZ, 2019).

Neste capítulo, é proposta então a análise de sinais de áudio de uma base de dados pública e de referência na literatura afim para avaliar se as características acústicas extraídas dos áudios são estatisticamente não-redundantes. É levantada a hipótese de que esse comportamento de *cluster* seja similar entre músicas de mesmo gênero musical e considera-se aqui uma metodologia de avaliação exploratória de seleção dessas características.

### 5.1 MATERIAIS E MÉTODOS

#### 5.1.1 Base de áudio

Foi utilizada a base de dados GTZAN (TZANETAKIS; COOK, 2002), por ser pública e bem conhecida na área de MIR. Esta base consiste em  $H = 1000$  trechos musicais (30 segundos cada) e contém 10 gêneros musicais com 100 músicas cada um, sendo esses: Blues, Classic, Country, Disco, Hiphop, Jazz, Metal, Pop, Reggae e Rock. A GTZAN é uma base de áudio largamente utilizada em Reconhecimento de Gênero Musical (MGR), mas apresenta algumas falhas que já foram apontadas por outros pesquisadores (STURM, 2014), tais como repetições, identificação incorreta e distorções.

### 5.1.2 Metodologia

Detalhando o método de seleção de características acústicas descrito na Seção 4.2.2, aqui é avaliado se todas as  $s = 12$  características acústicas de baixo nível extraídas das  $H$  músicas contidas na base de dados GTZAN são estatisticamente relevantes, a fim de prover uma melhor representação dos sinais de áudio por meio da Análise Fatorial (do inglês, *Factor Analysis - FA*) desses descritores sonoros.

Sejam, portanto, as variáveis  $\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_s$ . O modelo da análise fatorial é definido como (JOHNSON; WICHERN, 2007)

$$\begin{aligned} A_1 &= \mu_1 + l_{11}F_1 + l_{12}F_2 + \dots + l_{1R}F_R + \varepsilon_1 \\ A_2 &= \mu_2 + l_{21}F_1 + l_{22}F_2 + \dots + l_{2R}F_R + \varepsilon_2 \\ &\vdots \\ A_s &= \mu_s + l_{s1}F_1 + l_{s2}F_2 + \dots + l_{sR}F_R + \varepsilon_s \end{aligned} \quad (9)$$

ou, em notação matricial,

$$\mathbf{A} = \boldsymbol{\mu} + \mathbf{L}\mathbf{F} + \boldsymbol{\varepsilon} \quad (10)$$

sendo  $\boldsymbol{\mu}$  o vetor médio de  $\mathbf{A}$ ,  $\mathbf{L}$  a matriz de carga fatorial,  $\mathbf{F}$  o vetor de fatores comuns e  $\boldsymbol{\varepsilon}$  o vetor de erros.

Para estimar os parâmetros deste modelo, é utilizado o método das componentes principais (JOHNSON; WICHERN, 2007), pois a sua solução de decomposição espectral simplifica a questão de quantos fatores devem ser retidos no FA (RIBEIRO; THOMAZ, 2019), convencionalmente escolhendo os fatores cujos autovalores sejam maiores do que 1 para determinar o número adequado de fatores utilizados na análise (JOHNSON; WICHERN, 2007).

Idealmente, é desejado um padrão de resposta em que cada fator apresente alta carga fatorial para um determinado *cluster* de características acústicas e nos demais fatores as características desse *cluster* apresentem baixa carga fatorial. Portanto, os  $\mathbf{F} = [\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_R]$  fatores podem substituir as  $s$  variáveis iniciais, com  $R$  fatores comuns rotacionados. O propósito da rotação é rearranjar os fatores de tal forma que tenham uma relação mais simples e interpretável com os padrões implícitos nos dados (KAISER, 1958). O tipo de rotação aplicada pode permitir que os fatores se correlacionem ou não. Assim, esses se diferenciam entre rotações ortogonais ou oblíquas (JOHNSON; WICHERN, 2007). Neste caso é aplicado a rotação *Varimax*, que é a mais comum utilizada, sendo uma rotação ortogonal. Seu uso neste trabalho é explicado por

forçar os fatores a não se correlacionarem, simplificando a análise. Desta forma, as associações entre as características acústicas serão mais significativas em termos da variância dos dados (KAISER, 1958; JOHNSON; WICHERN, 2007).

De forma análoga ao capítulo anterior, os sinais de áudio são processados utilizando a toolbox MIRtoolbox (versão 1.6.1) no ambiente Matlab 2015a, decompondo o sinal em janelas de 50 ms de duração com 50% de sobreposição.

Seja, portanto, a matriz  $\mathbf{A}_g$  de cada gênero musical  $g$  representado na base de dados GTZAN descrita como:

$$\mathbf{A}_g = \begin{bmatrix} \mathbf{a}_{1,1} & \mathbf{a}_{1,2} & \dots & \mathbf{a}_{1,s} \\ \mathbf{a}_{2,1} & \mathbf{a}_{2,2} & \dots & \mathbf{a}_{2,s} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{a}_{H,1} & \mathbf{a}_{H,2} & \dots & \mathbf{a}_{H,s} \end{bmatrix}, \quad (11)$$

em que  $H$  é a música analisada e  $s$  a característica acústica extraída do sinal de áudio, tal que

$$\mathbf{a}_{H,s} = \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{J1} \end{bmatrix}, \quad (12)$$

em que  $J$  é o número total de janelas extraídas de cada característica acústica  $s$  e de cada música analisada  $H$ .

Dessa forma, a análise fatorial é aplicada na matriz  $\mathbf{A}_g$ , resultando na matriz  $\mathbf{F}_g$  que contém as cargas fatoriais das  $s$  características acústicas extraídas dos áudios, para os  $R$  fatores comuns rotacionados. Isso é feito para cada um dos  $g$  gêneros musicais. Ou seja, serão analisados os dados contidos no espaço das variáveis para se avaliar os comportamentos de *cluster* existentes entre os descritores sonoros.

Supõe-se que haja diferenças entre gêneros nos *clusters* resultantes da dispersão das cargas fatoriais de cada descritor sonoro. Deseja-se, portanto, avaliar também se os *clusters* gerados pela FA de cada gênero musical são similares, ou seja, determinar o grau de associação entre esses agrupamentos. Isso é feito aplicando a correlação de Pearson entre os  $R$  fatores comuns rotacionados para cada gênero musical da matriz  $\mathbf{F}_g$ .

O algoritmo  $k$ -means é empregado para particionar os dados em  $k$  conjuntos ( $k \in \{2, 3, \dots, 6\}$  aqui). Portanto, seja  $\mathbf{F}_g = [\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_R]$  os dados no espaço  $R$ -dimensional

das variáveis e seja  $C_i = C_1, C_2, \dots, C_k$  os agrupamentos em  $k$  *clusters*. Duas métricas serão calculadas aqui para determinar o valor ótimo de  $k$  e avaliar o número adequado de  $R$  fatores extraídos na FA, sendo essas: Silhouette e Davies-Boulding índices (ROUSSEEUW, 1987; DAVIES; BOULDIN, 1979).

### 5.1.2.1 Silhouette (Si)

O critério Silhouette (Si) (ROUSSEEUW, 1987) calcula-se a medida da similaridade entre as observações dentro de seu próprio *cluster* comparada com os outros *clusters* existentes. Assumimos que os dados já foram agrupados em  $k$  *clusters* pela técnica  $k$ -means. Assim, o critério Silhouette é definido como

$$S_i = \frac{(b_i - o_i)}{\max(b_i, o_i)}, \quad (13)$$

em que  $o_i$  é a distância média do ponto  $i$  aos outros pontos dentro de  $C_o$ , e  $b_i$  é a distância média mínima entre o ponto  $i$  e os pontos nos diferentes agrupamentos  $C_k$  para  $k \neq o$ , minimizado entre *clusters*.

Esta métrica quantifica quão bem o objeto  $i$  foi classificado. A solução ótima de agrupamento apresenta o maior valor do índice Silhouette.

### 5.1.2.2 Davies-Bouldin (DB)

O critério Davies-Bouldin (DB) (DAVIES; BOULDIN, 1979) é uma medida para avaliar algoritmos de agrupamentos baseada na razão das distâncias intra-cluster e inter-cluster. Novamente, assumimos que os dados já foram agrupados em  $k$  *clusters* pela técnica  $k$ -means. Assim, o critério DB é definido como

$$DB = \frac{1}{k} \sum_{i=1}^k D_i, \quad (14)$$

em que  $D_i$  é equivalente ao máximo de  $D_{i,j}$  para  $i \neq j$ , sendo

$$D_{i,j} = \frac{(\bar{d}_i + \bar{d}_j)}{d_{i,j}}, \quad (15)$$

tal que  $\bar{d}_i$  é a distância intra-cluster, i.e. a distância média entre cada ponto no *cluster*  $i$  ao seu centroide,  $\bar{d}_j$  a distância média entre cada ponto no *cluster*  $j$  ao seu centroide, e  $d_{i,j}$  a distância Euclidiana entre os centroides dos *clusters*  $i$  e  $j$ .

A solução ótima de agrupamento apresenta o menor valor do índice Davies-Bouldin.

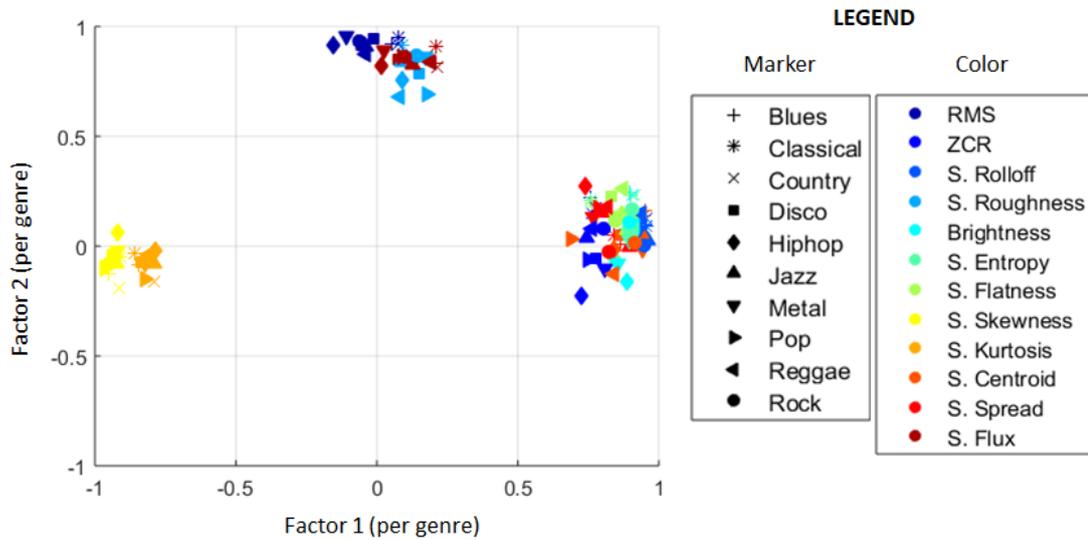
## 5.2 RESULTADOS

Os áudios foram analisados por meio da análise fatorial mencionada acima para cada gênero musical  $g$ . Para a maioria dos gêneros musicais, apenas 2 fatores apresentavam autovalor acima de 1, exceto os gêneros Metal e Pop, que apresentaram 3 fatores com autovalores acima de 1. Assim, inicialmente, foi definida a utilização de apenas 2 fatores para se analisar como as características acústicas extraídas dos áudios se comportavam entre os diferentes gêneros musicais. A Figura 27 apresenta os gráficos de dispersão dos 10 gêneros musicais existentes na base de áudio GTZAN com as cargas fatoriais das características acústicas extraídas nos 2 fatores gerados pela FA. É possível observar que mesmo para os diferentes gêneros musicais, as características acústicas se agrupam similarmente. Sendo assim, identificamos os *clusters* entre as seguintes características acústicas para todos os gêneros: Cluster 1 - 1 (RMS), 4 (Spectral Roughness) e 12 (Spectral Flux); Cluster 2 - 8 (Spectral Skewness) e 9 (Spectral Kurtosis); Cluster 3 - 2 (ZCR), 3 (Spectral Rolloff), 5 (Brightness), 6 (Spectral Entropy), 7 (Spectral Flatness), 10 (Spectral Centroid) e 11 (Spectral Spread).

Para avaliar o número adequado de fatores a serem empregados e a fim de determinar o número ótimo de *clusters*, os índices Silhouette (Figura 28) e Davies-Bouldin (Figura 30) foram utilizados, variando o número de  $k$ , considerando  $k \in \{2, 3, \dots, 6\}$ . Estes resultados mostram que para ambas as métricas, os melhores valores foram apresentados quando utilizado 2 fatores ( $R = 2$ ). Nessa condição, para todos os gêneros,  $k = 3$  apresenta o número ótimo de *clusters*. Para outras condições ( $R \neq 2$ ), o número de  $k$  variava significativamente entre gêneros. É possível ver a variação do número de  $k$  *clusters* nas Figuras 29 e 31.

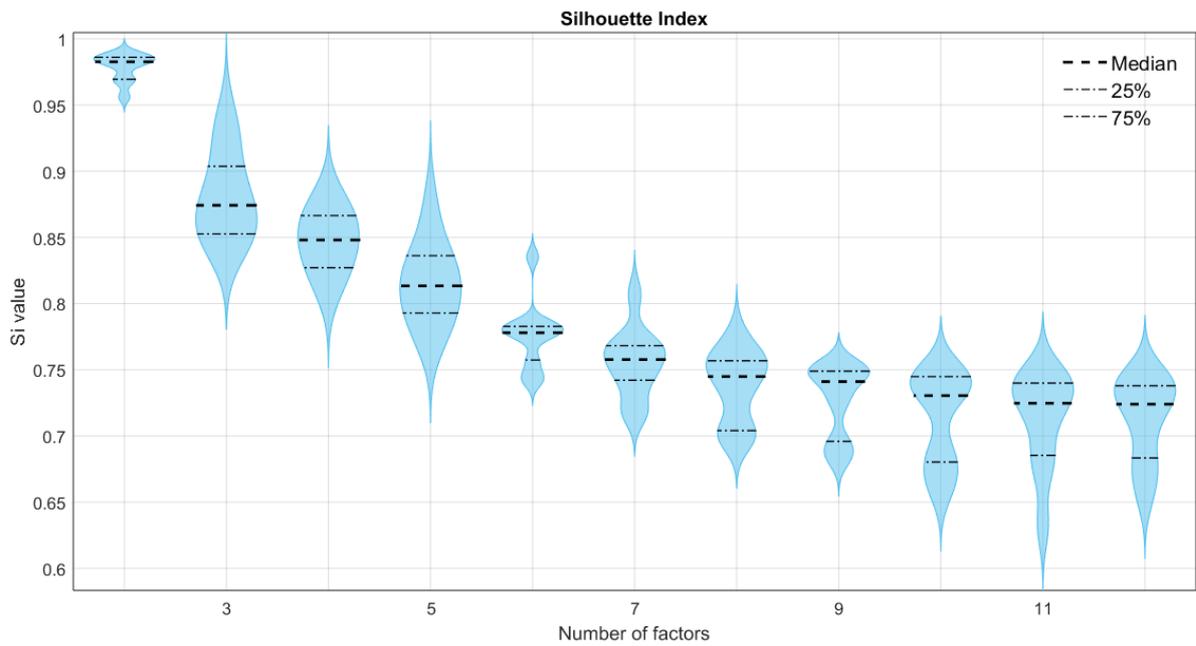
A matriz de correlação, para determinar o grau de associação entre as características acústicas dos gêneros musicais contidos na base de áudio GTZAN, é apresentada na Figura 32 que mostra a matriz de correlação do Fator 1 da FA, e na Figura 33 que mostra a matriz de correlação para o Fator 2 da FA. É possível ver que, utilizando  $R = 2$  como o número de fatores determinado pelos índices Silhouette e Davis-Bouldin, as características acústicas são

Figura 27 – Carga fatorial das características acústicas extraídas dos 10 diferentes gêneros musicais contidos na base de áudio GTZAN, sendo esses: Blues, Classical, Country, Disco, Hiphop, Jazz, Metal, Pop, Reggae, Rock.



Fonte: Autor.

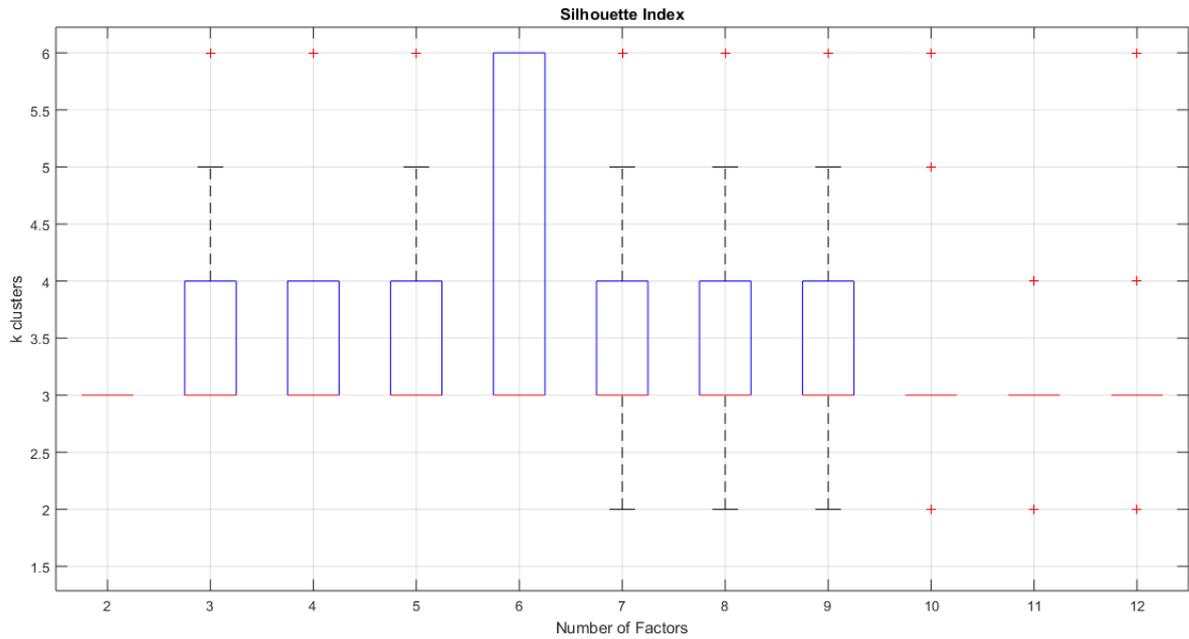
Figura 28 – Distribuição do índice Silhouette variando o número de fatores de 2 a 12 para os 10 gêneros musicais contidos na base de áudio GTZAN.



Fonte: Autor.

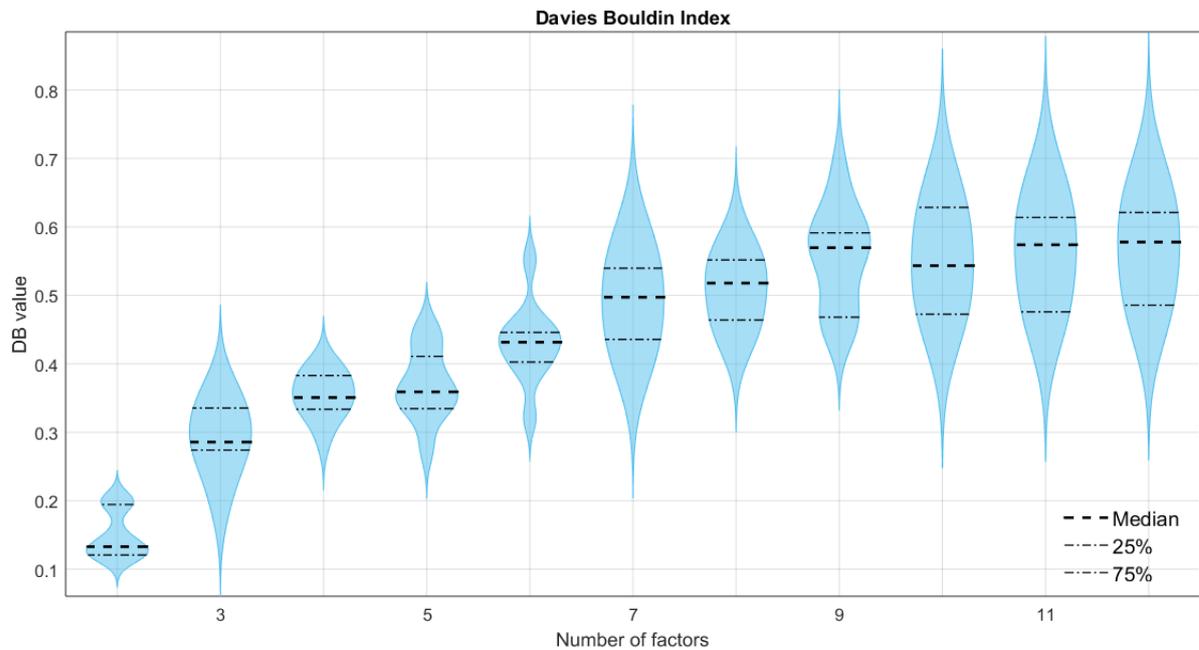
agrupadas de maneira similar, apresentando alta correlação entre os diferentes gêneros musicais para os dois fatores.

Figura 29 – Número de  $k$  identificados utilizando o índice Silhouette, variando o número de fatores de 2 a 12 para os 10 gêneros musicais contidos na base de áudio GTZAN.



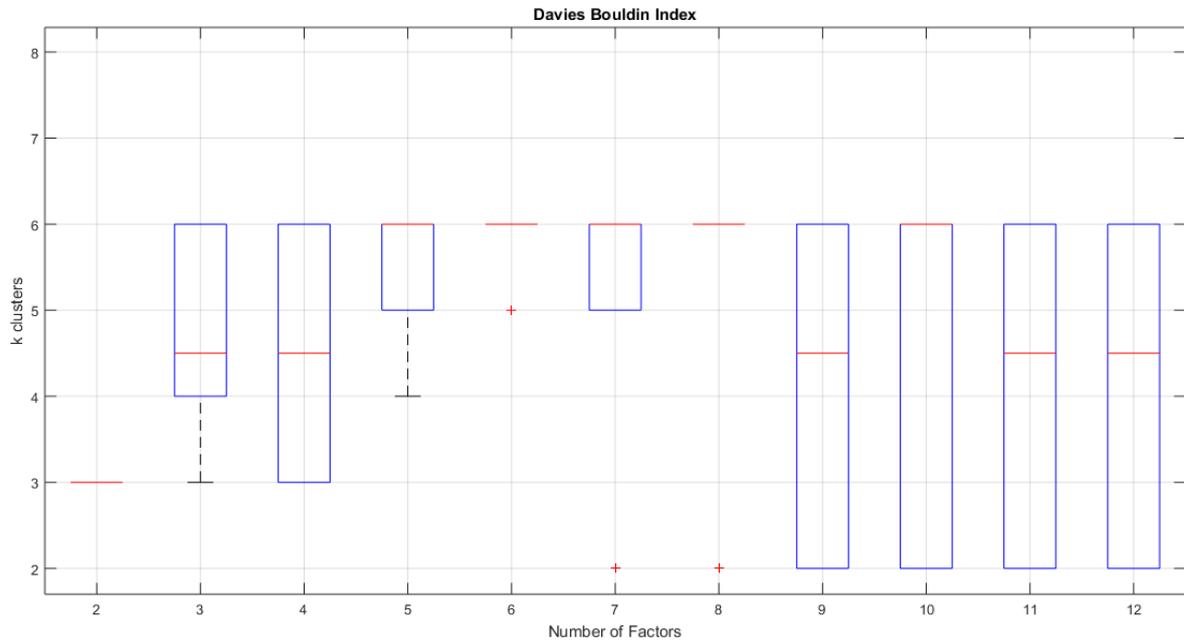
Fonte: Autor.

Figura 30 – Distribuição do índice Davis-Bouldin variando o número de fatores de 2 a 12 para os 10 gêneros musicais contidos na base de áudio GTZAN.



Fonte: Autor.

Figura 31 – Número de  $k$  identificados utilizando o índice Davis-Bouldin, variando o número de fatores de 2 a 12 para os 10 gêneros musicais contidos na base de áudio GTZAN.



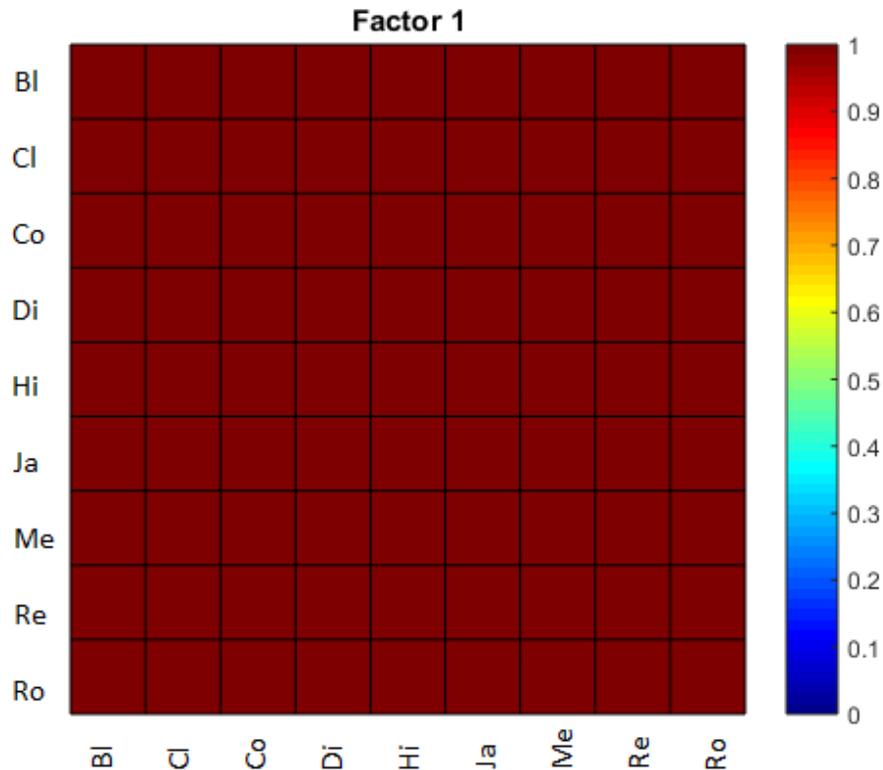
Fonte: Autor.

Similar ao que acontece na Figura 27, a Figura 34 apresenta os protótipos das cargas fatoriais das características acústicas considerando todos os áudios da base de dados GTZAN. Assim, usando 2 fatores, é possível observar os mesmos *clusters* encontrados anteriormente, sendo: Cluster 1 - 1 (RMS), 4 (Spectral Roughness) and 12 (Spectral Flux); Cluster 2 - 8 (Spectral Skewness) and 9 (Spectral Kurtosis); Cluster 3 - 2 (ZCR), 3 (Spectral Rolloff), 5 (Brightness), 6 (Spectral Entropy), 7 (Spectral Flatness), 10 (Spectral Centroid) e 11 (Spectral Spread).

### 5.3 DISCUSSÃO

Neste capítulo, foi aplicada a análise estatística multivariada para entender como se relacionam o conjunto de descritores sonoros utilizados nesta tese. Nossos principais resultados mostram que independente do gênero musical, o mesmo comportamento de agrupamento acontece para as características acústicas utilizadas nesse trabalho. Isso pode ser observado na Figura 32 e 33, em que todas as características acústicas apresentam altos valores de correlação entre diferentes gêneros, significando que os mesmos *clusters* são encontrados.

Figura 32 – Matriz de correlação do Fator 1 entre os 10 gêneros musicais contidos na base de áudio GTZAN, utilizando 2 fatores extraídos da FA.

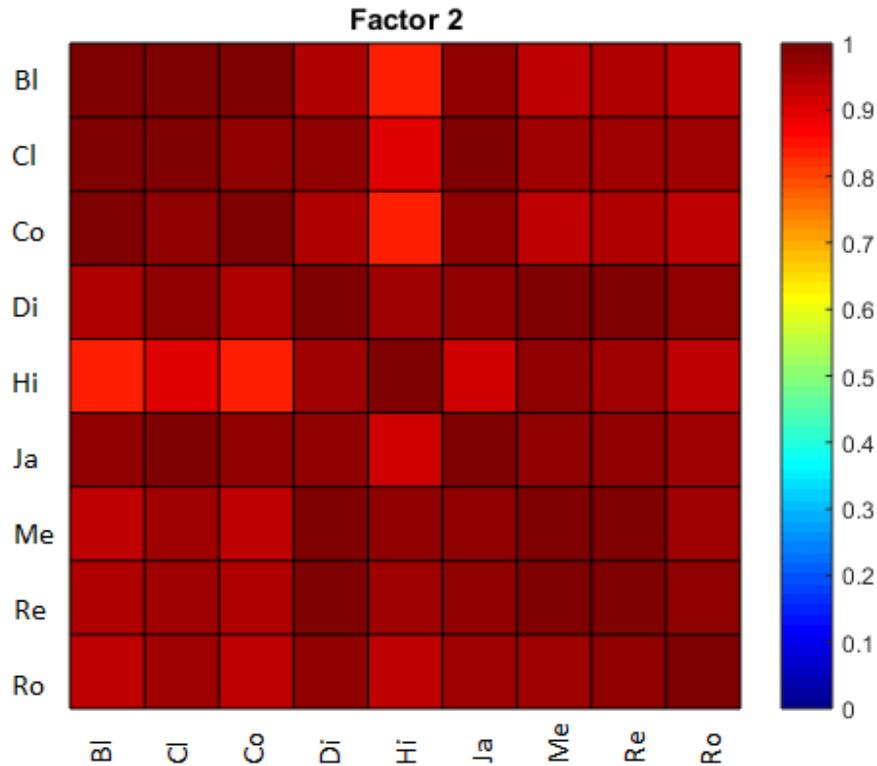


Fonte: Autor.

O objetivo de utilizar a FA é revelar a associação entre um conjunto de características, selecionando as mais relevantes. Isso é importante porque a seleção de características afeta fortemente análises posteriores como a classificação ou a identificação de instantes significativos nos áudios capazes de gerar respostas neurais relevantes durante tarefas auditivas. Isso é um problema de reconhecimento de padrões que estima as funções de densidade em espaços altamente dimensionais, dividindo-os em regiões, determinando um comportamento de *cluster* entre o conjuntos de características extraídas dos sinais de áudio (JOHNSON; WICHERN, 2007).

A princípio foi levantada a hipótese de que os gêneros musicais apresentariam diferentes comportamentos de *clusters* para o conjunto de características acústicas extraídas dos áudios. Nossa análise mostra que o mesmo *cluster* acontece, independente do gênero. Dada a natureza de cada característica acústica, é razoável que algumas dessas se agrupem, como no *Cluster 3*, em que todas as características são relacionadas com propriedades da tonalidade e ruído, ou o menor *cluster*, em que Spectral Skewness e Spectral Kurtosis estão relacionados

Figura 33 – Matriz de correlação do Fator 2 entre os 10 gêneros musicais contidos na base de áudio GTZAN, utilizando 2 fatores extraídos da FA.

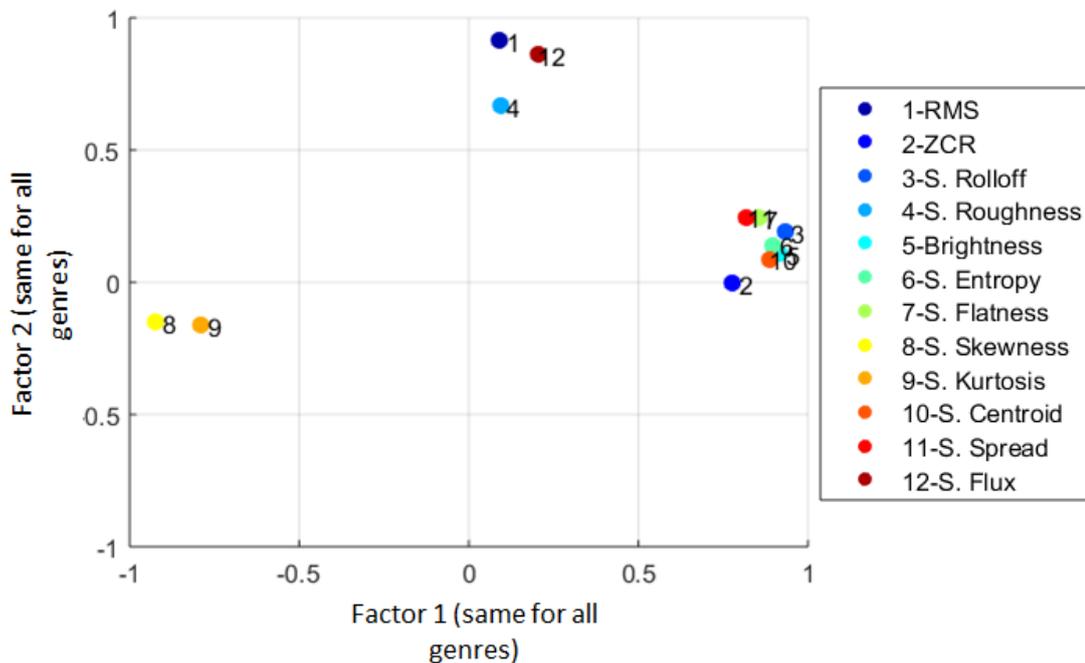


Fonte: Autor.

com a forma espectral da distribuição dos dados. As características Spectral Flux e Spectral Roughness estão associadas com a sensação de dissonância no áudio e estão agrupadas também. Isso acontece especificamente quando são utilizados 2 fatores na FA. As Figuras 28 e 30 claramente demonstram que  $R = 2$  apresenta a melhor relação de agrupamento, confirmando que  $k = 3$  é o número ótimo de agrupamentos para o arranjo dessas características acústicas. Aparentemente o primeiro e segundo fatores são melhores para explicar a similaridade entre as características acústicas extraídas para todos os gêneros musicais. No entanto, existem diferenças intra-*clusters* entre os gêneros musicais, embora o padrão inter-*cluster* permaneça igual.

Uma abordagem semelhante à apresentada neste trabalho é a de Alluri et al. (2012). Em seu experimento, utilizando fMRI, foi realizada a Análise Fatorial em um conjunto de 25 características acústicas, compreendendo características de baixo e alto nível, utilizando a música clássica *Adios Nonino* de Astor Piazzolla. No entanto, para se adequar a frequência de amostragem dos sinais de fMRI, foi realizada a redução da taxa de amostragem (*down-sampled*) das séries temporais que resultam das características acústicas de tal forma que o estímulo usado

Figura 34 – Protótipo das cargas fatoriais das características acústicas extraídas de todos os áudios contidos na base de áudio GTZAN.



Fonte: Autor.

(que possuía 7 minutos e 42 segundos de duração) resultou em um vetor de 231 amostras. Esta redução de amostragem impacta diretamente na informação extraída das características, contrastando com a resolução temporal que encontramos no EEG. Mesmo com esta limitação, Alluri et al. (2012) computaram o PCA com rotação *varimax*, retendo 9 componentes principais, encontrando agrupamentos entre as características acústicas extraídas. As características de alto nível não formaram agrupamentos, mas cada uma correspondeu a um PC específico. Já as características de baixo nível formaram agrupamentos semelhantes aos encontrados neste trabalho, sendo esses: (PC2) S. Centroid, S. Rolloff e ZCR (correspondendo ao aspecto do brilho do sinal); (PC3) S. Spread e S. Flatness (correspondendo ao aspecto da complexidade do timbre); (PC9) Roughness e S. Flux (correspondendo ao aspecto “*Activity*”); e RMS separado das outras características (correspondendo ao aspecto da intensidade). Vemos, assim, uma conformidade entre os resultados obtidos por Alluri et al. (2012) e os encontrados neste trabalho, e aqui demonstrando uma generalização destes resultados para gêneros distintos de músicas, demonstrando matematicamente e experimentalmente que essas características de baixo nível são, de fato, redundantes

Finalmente, essa abordagem, proposta para seleção de características, pode ser utilizada em problemas diversos, em que o número de características é maior. Assim, neste capítulo

foi demonstrado que é possível selecionar características acústicas que melhor representem as músicas apresentadas como estímulo em tarefas auditivas, a fim de investigar como o cérebro processa a música, dada a redundância estatística encontrada na FA, independentemente da própria música e do seu gênero.

#### 5.4 CONSIDERAÇÕES COMPLEMENTARES: PADRÕES COGNITIVOS E SONOROS

Tendo em vista os resultados obtidos no capítulo anterior e neste, é possível determinar dois critérios para a seleção das características acústicas que podem ser utilizados na análise dos sinais de EEG, posterior à análise dos sinais de áudio.

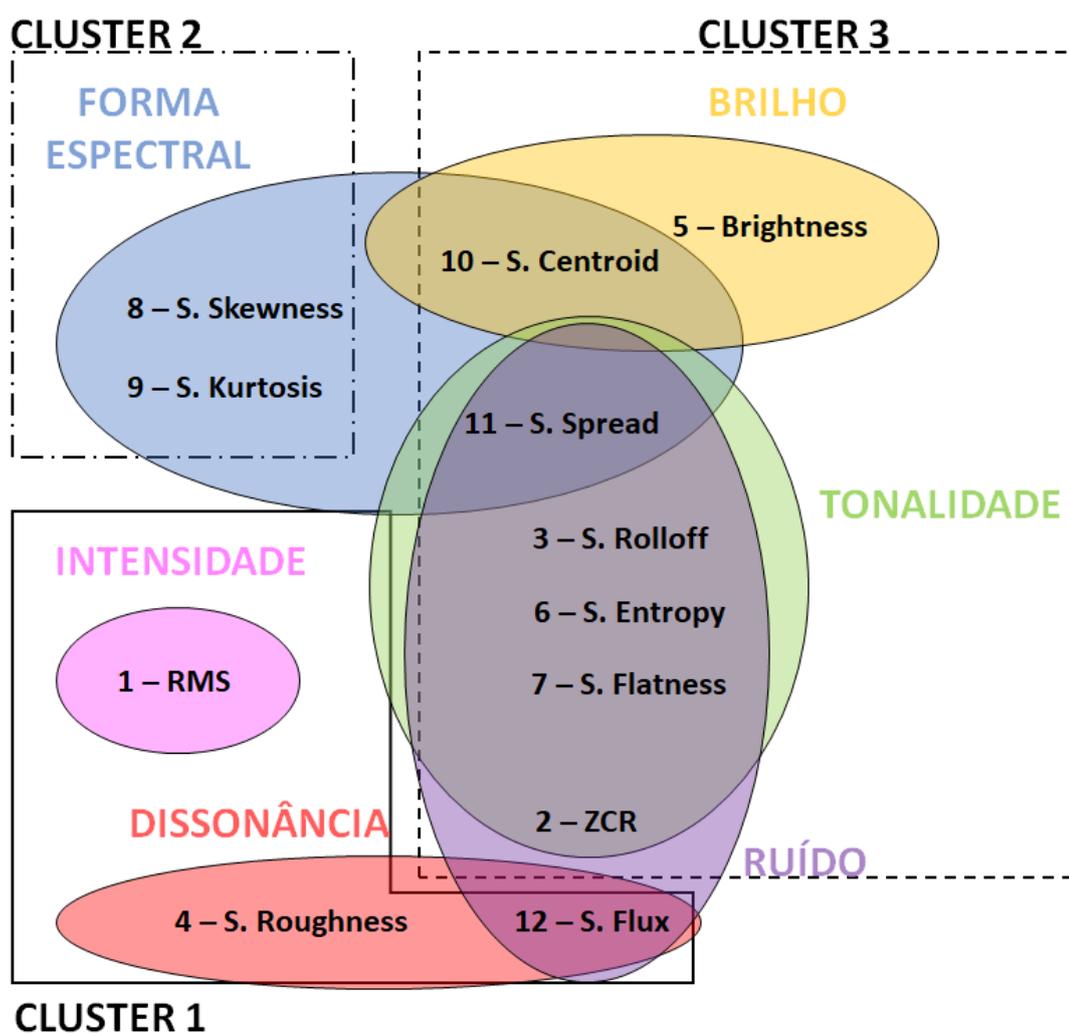
O primeiro critério é o utilizado no capítulo anterior, em que são selecionadas as características acústicas com maior carga fatorial (em módulo) dentro de cada agrupamento identificado no espaço das variáveis obtidos pelo FA. Ou seja, caso se esteja interessado em escolher as características mais relevantes, deve-se utilizar esse primeiro critério. O segundo critério seria concatenar os *triggers* das características do mesmo agrupamento, aumentando a representatividade e a quantidade de amostras para a análise dos sinais de EEG subsequentes, conforme proposto em Ribeiro e Thomaz (2019). Considerando que foi demonstrado que as características agrupadas são estatisticamente redundantes, é razoável dizer que a informação contida nessas características são similares e, portanto, estaríamos extraíndo informações cognitivas resultantes de um mesmo tipo de estímulo. Ou seja, caso esteja interessado em ter uma maior representação dos dados, ou mais informações para analisar os sinais de EEG, deve-se utilizar esse segundo critério.

Utilizando o primeiro critério, obteríamos sempre a seleção das mesmas características acústicas por grupos (ver as Tabelas 5-14 no Apêndice B), sendo essas: RMS, Spectral Skewness e Spectral Rolloff. Com exceção do gênero musical “Metal”, em que a característica Spectral Rolloff é substituída pela característica Spectral Centroid. Assim, seria possível dizer que para este critério, estas características seriam sempre as representantes para os agrupamentos, não havendo a necessidade de extrair as outras. No entanto, observa-se pela Figura 27 que embora os agrupamentos sejam iguais, há diferenças *intra-clusters* para cada gênero musical que podem ser relevantes. Portanto, é importante levar essa questão em consideração.

Levando em consideração a descrição das características acústicas (ver Apêndice A) e os resultados dos agrupamentos obtidos neste capítulo, é possível discutir de maneira inédita

a relação entre a forma como estas características são descritas pela literatura e como essas características são agrupadas estatisticamente, como é apresentado na Figura 35.

Figura 35 – Relação entre os 12 descritores sonoros de baixo nível utilizados neste trabalho e os 3 clusters encontrados pela FA.



Fonte: Autor.

Interessante observar que os agrupamentos propostos pela FA (Figura 19) separam algumas características em dois *clusters* que estão intimamente relacionados com atributos da percepção auditiva (*cluster 1* e *cluster 3*), e um que está relacionado com medidas estatísticas (*cluster 2*) que são difíceis de encontrar correspondência com estas representações perceptuais dos áudios. A priori, espera-se que estes descritores sonoros com representações perceptuais forneçam melhores informações sobre o processamento neural da música e melhor discriminação dos grupos amostrados. No entanto, os resultados do capítulo anterior demonstram que a característica que melhor discriminou os grupos e trouxe melhores descrições dos padrões de

ativação corticais foi a característica Spectral Kurtosis, que é um descritor estatístico. Ou seja, parece que a informação que usamos para processar as características acústicas das músicas é uma informação implicitamente estatística.

## 6 CONCLUSÃO

A compreensão do funcionamento do cérebro humano requer o uso de estímulos biologicamente realísticos. Este princípio foi aplicado nesta tese para estudar o processamento da música pelo cérebro, apresentando uma música completa durante a tarefa auditiva. É possível identificar padrões de ativações corticais que descrevem os grupos analisados, visualizando as mudanças nos potenciais elétricos ao longo dos hiperplanos discriminantes. Assim, de um lado do eixo vimos como se caracteriza a ativação neural para o conjunto de voluntários não-músicos e, conforme navegamos ao longo desta dimensão, podemos ver como esta ativação neural se altera, passando por uma região de transição entre classes, até chegar ao outro lado do eixo, caracterizando a ativação neural para o conjunto de voluntários músicos.

Vimos também que não há necessidade de se utilizar todas as características acústicas disponíveis na literatura afim para tal análise, devido à redundância de informação nessas características. Podemos, na verdade, utilizar esse achado para reduzir a dimensionalidade desses dados multidimensionais e aumentar a quantidade de amostras nas identificações dos *triggers*, representando melhor os sinais de interesse.

Duas questões importantes precisam ser levadas em consideração como trabalhos futuros. A primeira questão se refere à representação dos dados, ou seja, como formar a matriz de dados  $X_r$  para que haja um melhor aproveitamento da informação disponível. Potenciais abordagens, já mencionadas anteriormente, poderiam ser empregadas, como a sintetização da informação contida em cada época do sinal de EEG, por meio do método de Entropia da Correlação (ROCHA et al., 2014), por exemplo, ou poderiam ser analisadas por meio de uma perspectiva de Análise Multilinear, considerando tempo, eletrodos e sujeitos como um tensor de terceira ordem (LU; PLATANIOTIS; VENETSANOPOULOS, 2013). A segunda questão se refere ao número limitado de amostras disponíveis para a análise desses dados. Este problema pode ser abordado utilizando versões regularizadas do LDA, como o *Maximum uncertainty Linear Discriminant Analysis* (MLDA) (THOMAZ et al., 2007; SATO et al., 2008), ou outros classificadores como a Máquina de Vetores Suporte (em inglês, *Support Vector Machines* - SVM) (DAVATZIKOS, 2004). Além disso, tendo em vista a variação existente dentro da classe de músicos (intra-classe), seria interessante tentar expandir essa classe e então propor um sistema de classificação multi-classes, considerando os diferentes tipos de músicos existentes. Para isso, seria necessário aumentar o número de voluntários, representando cada classe de forma menos esparsa.

Destaque deve ser dado ao fato de que foi possível separar os voluntários, mesmo considerando que os músicos são amadores. Esse resultado corrobora com o de Tervaniemi (2006), quanto a distinção de não-músicos e músicos amadores, e nos faz pensar no impacto que a música causa, afinal, mesmo uma pequena prática musical é capaz de mudar o processamento neural de tal forma que é possível distingui-los dos não-músicos.

Acredita-se que a metodologia proposta possa ser utilizada para se investigar preferências musicais baseadas nas ativações cerebrais durante tarefas auditivas. Desta forma, seria identificado quais os gêneros musicais os voluntários gostam e não-gostam, podendo relacionar esta informação com traços de suas personalidades (RENTFROW; GOSLING, 2003; GREENBERG et al., 2015). Outra abordagem seria a caracterização e discriminação dos padrões de ativações corticais de músicos e não-músicos inerentes ao processo de envelhecimento. Acredita-se ser possível também utilizar essa metodologia de seleção de *triggers* para gerar estímulos sonoros capazes de gerar respostas neurais significativas. Ou seja, criar músicas artificiais baseadas nesses trechos identificados pelos *triggers*. Esta metodologia de seleção de características seria importante nesse caso, para determinar quais características deveriam ser utilizadas.

Por fim, faz-se importante mencionar que os métodos desta tese são baseados em sinais de EEG, evidenciando somente padrões cognitivos corticais. O uso de outros tipos de sinais, tais como fMRI, podem aumentar a compreensão do funcionamento do cérebro humano em nível subcortical. No entanto, em especial com relação a experimentos de fMRI, os experimentos seriam menos ecológicos que os propostos aqui.

## REFERÊNCIAS

- ABRAMS, Daniel A. et al. Inter-subject synchronization of brain response during natural music listening. **European Journal of Neuroscience**, v. 37, n. 9, p. 1458–1469, 2013. DOI: 10.1111/ejn.12173.
- ALAIN, Claude et al. Turning down the noise: The benefit of musical training on the aging auditory brain. **Hearing Research**, v. 308, p. 162–173, 2014. Music: A window into the hearing brain. DOI: 10.1016/j.heares.2013.06.008.
- ALLURI, Vinoo et al. From Vivaldi to Beatles and back: Predicting lateralized brain responses to music. **NeuroImage**, v. 83, p. 627–636, dez. 2013. DOI: 10.1016/j.neuroimage.2013.06.064.
- ALLURI, Vinoo et al. Large-scale brain networks emerge from dynamics processing of musical timbre, key and rhythm. **NeuroImage**, v. 59, n. 4, p. 3677–3689, 2012. DOI: 10.1016/j.neuroimage.2011.11.019.
- ALTENMÜLLER, Eckart; SCHLAUG, Gottfried. Chapter 12 - Apollo's gift: new aspects of neurologic music therapy. In: ALTENMÜLLER, Eckart; FINGER, Stanley; BOLLER, François (Ed.). **Music, Neurology, and Neuroscience: Evolution, the Musical Brain, Medical Conditions, and Therapies**. [S.l.]: Elsevier, 2015. v. 217. (Progress in Brain Research). P. 237–252. DOI: 10.1016/bs.pbr.2014.11.029.
- ANGULO-PERKINS, Arafat et al. Music listening engages specific cortical regions within the temporal lobes: differences between musicians and non-musicians. **Cortex**, v. 59, p. 126–137, out. 2014. DOI: 10.1016/j.cortex.2014.07.013.
- ANTONI, Jérôme. The spectral kurtosis: a useful tool for characterizing non-stationary signals. **Mechanical systems and signal processing**, v. 20, n. 2, p. 282–307, 2006. DOI: 10.1016/j.ymsp.2004.09.001.
- BAIRD, Ameer; SAMSON, Séverine. Chapter 11 - Music and dementia. In: ALTENMÜLLER, Eckart; FINGER, Stanley; BOLLER, François (Ed.). **Music, Neurology, and Neuroscience: Evolution, the Musical Brain, Medical Conditions, and Therapies**. [S.l.]: Elsevier, 2015. v. 217. (Progress in Brain Research). P. 207–235. DOI: 10.1016/bs.pbr.2014.11.028.
- BALARDIN, Joana et al. Imaging brain function with functional-near-infrared spectroscopy in unconstrained environments. **Frontiers in human neuroscience**, v. 11, p. 258, 2017. DOI: 10.3389/fnhum.2017.00258.
- BANGERT, Marc; SCHLAUG, Gottfried. Specialization of the specialized in features of external human brain morphology. **European Journal of Neuroscience**, v. 24, n. 6, p. 1832–1834, 2006. DOI: 10.1111/j.1460-9568.2006.05031.x.
- BARRETT, Karen et al. Art and science: how musical training shapes the brain. **Frontiers in Psychology**, v. 4, p. 713, 2013. DOI: 10.3389/fpsyg.2013.00713.
- BARTOLOTTI, Nancy; LAZAROV, Orly. Chapter 7 - Lifestyle and Alzheimer's Disease: The Role of Environmental Factors in Disease Development. In: LAZAROV, Orly; TESCO, Giuseppina (Ed.). **Genes, Environment and Alzheimer's Disease**. San Diego: Academic Press, 2016. P. 197–237. DOI: 10.1016/B978-0-12-802851-3.00007-3.
- BEAR, Mark F.; CONNORS, Barry W.; PARADISO, Michael A. **Neuroscience: exploring the brain**. 4th. [S.l.]: Lippincott Williams & Wilkins, 2015.

- BENGTSSON, S.L. et al. Extensive piano practicing has regionally specific effects on white matter development. **Nat Neurosci**, v. 8, n. 8, p. 1148–1150, ago. 2005. DOI: 10.1038/nn1516.
- BHERER, Louis. Cognitive plasticity in older adults: effects of cognitive training and physical exercise. **Annals of the New York Academy of Sciences**, v. 1337, n. 1, p. 1–6, 2015. DOI: 10.1111/nyas.12682.
- BIDELMAN, Gavin M.; ALAIN, Claude. Musical training orchestrates coordinated neuroplasticity in auditory brainstem and cortex to counteract age-related declines in categorical vowel perception. **Journal of neuroscience**, v. 35, n. 3, p. 1240–1249, 2015. DOI: 10.1523/JNEUROSCI.3292-14.2015.
- BOSTANOV, Vladimir. Multivariate assessment of event-related potentials with the t-CWT method. **BMC Neurosci**, v. 16, p. 73, 2015. DOI: 10.1186/s12868-015-0185-z.
- BROWN, Rachel M.; ZATORRE, Robert J.; PENHUNE, Virginia B. Chapter 4 - Expert music performance: cognitive, neural, and developmental bases. In: ALTENMÜLLER, Eckart; FINGER, Stanley; BOLLER, François (Ed.). **Music, Neurology, and Neuroscience: Evolution, the Musical Brain, Medical Conditions, and Therapies**. [S.l.]: Elsevier, 2015. v. 217. (Progress in Brain Research). P. 57–86. DOI: 10.1016/bs.pbr.2014.11.021.
- BUZSÁKI, György. **Rhythms of the brain**. 1st. New York: Oxford University, 2006. ISBN 0-19-530106-4.
- BUZSÁKI, György; ANASTASSIOU, Costas A.; KOCH, Christof. The origin of extracellular field and currents: EEG, ECoG, LFP and spikes. **Nature Review**, v. 13, p. 407–420, 2012. DOI: 10.1038/nrn3241.
- CARPENTIER, Sarah M.; MORENO, Sylvain; MCINTOSH, Anthony R. Short-term Music Training Enhances Complex, Distributed Neural Communication during Music and Linguistic Tasks. **Journal of Cognitive Neuroscience**, v. 28, n. 10, p. 1603–1612, 2016. DOI: 10.1162/jocn\_a\_00988.
- CHEN, Joyce L.; PENHUNE, Virginia B.; ZATORRE, Robert J. Moving on Time: Brain Network for Auditory-Motor Synchronization is Modulated by Rhythm Complexity and Musical Training. **Journal of Cognitive Neuroscience**, v. 20, n. 2, p. 226–239, 2008. DOI: 10.1162/jocn.2008.20018.
- COOTES, T. F.; EDWARDS, G. J.; TAYLOR, C. J. Active appearance models. In: BURKHARDT, Hans; NEUMANN, Bernd (Ed.). **Computer Vision - ECCV'98**. Berlin, Heidelberg: Springer Berlin Heidelberg, 1998. P. 484–498. ISBN 978-3-540-69235-5.
- DAS, Koel; GIESBRECHT, Barry; ECKSTEIN, Miguel P. Predicting variations of perceptual performance across individuals from neural activity using pattern classifiers. **NeuroImage**, v. 51, n. 4, p. 1425–1437, 2010. DOI: 10.1016/j.neuroimage.2010.03.030.
- DAVATZIKOS, Christos. Why voxel-based morphometric analysis should be used with great caution when characterizing group differences. **NeuroImage**, v. 23, n. 1, p. 17–20, 2004. DOI: 10.1016/j.neuroimage.2004.05.010.
- DAVIES, D. L.; BOULDIN, D. W. A Cluster Separation Measure. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, PAMI-1, n. 2, p. 224–227, 1979. DOI: 10.1109/TPAMI.1979.4766909.
- DECARLO, Lawrence T. On the meaning and use of kurtosis. **Psychological methods**, v. 2, n. 3, p. 292–307, 1997. DOI: 10.1037/1082-989X.2.3.292.

- DELORME, Arnaud; MAKEIG, Scott. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. **Journal of Neuroscience Methods**, v. 134, n. 1, p. 9–21, 2004. DOI: 10.1016/j.jneumeth.2003.10.009.
- DUSHANOVA, Juliana; CHRISTOV, Mario. The effect of aging on EEG brain oscillations related to sensory and sensorimotor functions. **Advances in Medical Sciences**, v. 59, n. 1, p. 61–67, 2014. DOI: 10.1016/j.advms.2013.08.002.
- DWYER, Roger F. Use of the kurtosis statistic in the frequency domain as an aid in detecting random signals. **IEEE Journal of Oceanic Engineering**, v. 9, n. 2, p. 85–92, 1984. DOI: 10.1109/JOE.1984.1145602.
- EGGERMONT, Jos J. Chapter 9 - Music and the Brain. In: EGGERMONT, Jos J. (Ed.). **Noise and the Brain**. San Diego: Academic Press, 2014. P. 240–265. ISBN 978-0-12-415994-5. DOI: 10.1016/B978-0-12-415994-5.00009-9.
- ELMER, S.; HANGGI, J.; JANCKE, L. Interhemispheric transcallosal connectivity between the left and right planum temporale predicts musicianship, performance in temporal speech processing, and functional specialization. **Brain Struct Funct**, v. 221, n. 1, p. 331–344, 2016. DOI: 10.1007/s00429-014-0910-x.
- FERREIRA, Leonardo Antunes. Um estudo sobre os padrões de ativação corticais gerados durante a audição de músicas brasileiras. In: SICFEI - Simpósio de Iniciação Científica, Didática e de Ações Sociais de Extensão da FEI. São Bernardo do Campo, Brasil: [s.n.], 2019.
- FERREIRA, Leonardo Antunes; RIBEIRO, Estela; THOMAZ, Carlos Eduardo. A multivariate statistical analysis of EEG signals for differentiation of musicians and non-musicians. In: ARAÚJO, João Teixeira; SCHIAVONI, Flávio Luiz (Ed.). **17th Brazilian Symposium on Computer Music (SBCM)**. Sao Paulo, Brazil: Brazilian Computer Society (SBC), out. 2019. P. 80–85.
- FRISTON, K. J.; ASHBURNER, J. Generative and recognition models for neuroanatomy. **NeuroImage**, v. 23, n. 1, p. 21–24, 2004. DOI: 10.1016/j.neuroimage.2004.04.021.
- FUKUNAGA, Keinosuke. **Introduction to statistical pattern recognition**. 2nd. Boston: Academic Press, 1990. DOI: 10.1016/C2009-0-27872-X.
- GASER, Christian; SCHLAUG, Gottfried. Brain structure differ between musicians and nonmusicians. **Journal of neuroscience**, v. 23, n. 27, p. 9240–9245, out. 2003. DOI: 10.1523/JNEUROSCI.23-27-09240.2003.
- GASER, Christian; SCHLAUG, Gottfried. Gray matter differences between musicians and nonmusicians. **Ann N Y Acad Sci**, v. 999, n. 11, p. 514–517, 2003. DOI: 10.1196/annals.1284.062.
- GEBEL, B. et al. Instrument specific brain activation in sensorimotor and auditory representation in musicians. **NeuroImage**, v. 74, p. 37–44, 2013. DOI: 10.1016/j.neuroimage.2013.02.021.
- GREENBERG, David M. et al. Musical Preferences are Linked to Cognitive Styles. **PLOS ONE**, Public Library of Science, v. 10, n. 7, p. 1–22, jul. 2015. DOI: 10.1371/journal.pone.0131151.
- GREGORI, I. R. S.; SANCHES, I.; THOMAZ, C. E. Clutch Judder Classification and Prediction: A Multivariate Statistical Analysis Based on Torque Signals. **IEEE Transactions on Industrial Electronics**, v. 64, n. 5, p. 4287–4295, mai. 2017. ISSN 0278-0046. DOI: 10.1109/TIE.2016.2630666.

- GRIFFITHS, TIMOTHY D. The Neural Processing of Complex Sounds. **Annals of the New York Academy of Sciences**, v. 930, n. 1, p. 133–142, 2001. DOI: 10.1111/j.1749-6632.2001.tb05729.x.
- GROPPE, David M.; URBACH, Thomas P.; KUTAS, Marta. Mass univariate analysis of event-related brain potentials/fields I: A critical tutorial review. **Psychophysiology**, v. 48, n. 12, p. 1711–1725, 2011. DOI: 10.1111/j.1469-8986.2011.01273.x.
- HABECK, Christian et al. Brain Networks Associated with Cognitive Reserve in Healthy Young and Old Adults. **Cerebral Cortex**, v. 15, n. 4, p. 394–402, ago. 2004. DOI: 10.1093/cercor/bhh142.
- HO, Ming-Chung et al. Age-related changes of task-specific brain activity in normal aging. **Neuroscience letters**, v. 507, n. 1, p. 78–83, 2012. DOI: 10.1016/j.neulet.2011.11.057.
- HORNBY-TURNER, Yvonne Claire; PEEL, Nancye May; HUBBARD, Ruth Eleanor. Health assets in older age: a systematic review. **BMJ Open**, British Medical Journal Publishing Group, v. 7, n. 5, 2017. DOI: 10.1136/bmjopen-2016-013226.
- HUTCHINSON, S. et al. Cerebellar volume of musicians. **Cereb cortex**, v. 13, n. 9, p. 943–9, set. 2003. DOI: 10.1093/cercor/13.9.943.
- HYDE, Krista L. et al. Musical Training Shapes Structural Brain Development. **Journal of Neuroscience**, v. 29, n. 10, p. 3019–3025, 2009. DOI: 10.1523/JNEUROSCI.5118-08.2009.
- JOHNSON, Richard A.; WICHERN, Dean W. **Applied Multivariate Statistical Analysis**. 6th. [S.l.]: Pearson, 2007. ISBN 01301877151.
- KAISER, H. F. The varimax criterion for analytic rotation in factor analysis. **Psychometrika**, v. 23, p. 187–200, 1958. DOI: 10.1007/BF02289233.
- KAMEL, Nidal; MALIK, Aamir Saeed. **EEG/ERP analysis: methods and applications**. 1st. New York: CRC, 2014.
- KANDEL, Eric R. et al. **Principles of neural science**. 5th ed. [S.l.]: McGraw-Hill, 2012.
- KLEIN, Carina et al. The “silent” imprint of musical training. **Human brain mapping**, v. 37, n. 2, p. 536–546, 2016. DOI: 10.1002/hbm.23045.
- KNESS, Peter; SCHEDL, Markus. **Music Similarity and Retrieval: An introduction to audio and web based strategies**. 1st. [S.l.]: Springer, 2016. ISBN 9783662497227.
- KOELSCH, S.; SCHRÖGER, E.; TERVANIEMI, M. Superior pre-attentive auditory processing in musicians. **Neuroreports**, v. 10, n. 6, p. 1309–13, 1999. DOI: 10.1097/00001756-199904260-00029.
- KOELSCH, Stefan et al. Brain indices of music processing: nonmusicians are musical. **Journal of Cognitive Neuroscience**, v. 12, n. 3, p. 520–541, 2000. DOI: 10.1162/089892900562183.
- KOENEKE, S. et al. Long-term training affects cerebellar processing in skilled keyboard players. **NeuroReport**, v. 15, n. 8, p. 1279–1282, 2004. DOI: 10.1097/01.wnr.0000127463.10147.e7.
- KOK, Albert. Age-related changes in involuntary and voluntary attention as reflected in components of the event-related potential (ERP). **Biological Psychology**, v. 54, n. 1-3, p. 107–143, 2000. DOI: 10.1016/S0301-0511(00)00054-5.
- KRAUS, Nina; NICOL, Trent. Auditory Evoked Potentials. In: BINDER, Marc D.; HIROKAWA, Nobutaka; WINDHORST, Uwe (Ed.). **Encyclopedia of Neuroscience**. Berlin, Heidelberg: Springer, 2008. P. 214–218. DOI: 10.1007/978-3-642-72658-3\_4.

- KREBS, Claudia; WEIBERG, Joanne; AKESSON, Elizabeth. **Lippincotts illustrated reviews: neuroscience**. 1st ed. [S.l.]: Lippincott Williams e Wilkins, 2011.
- KROPOTOV, Juri. **Quantitative EEG, Event-Related Potentials and Neurotherapy**. 1st. [S.l.]: Academic Press, 2008.
- LARTILLOT, Oliver. **MIRtoolbox 1.6.1 Users Manual**. English. Versão Version 1.6.1. [S.l.], dez. 2014. 220 p. December 7, 2014.
- LEHMANN, D. From mapping to the analysis and interpretation of EEG/EP maps. In: K., Maurer (Ed.). **Topographic Brain Mapping of EEG and Evoked Potentials**. Berlin, Heidelberg: Springer, 1989. P. 53–75. DOI: 10.1007/978-3-642-72658-3\_4.
- LENT, Roberto. **Cem bilhões de neurônios: conceitos fundamentais de neurociência**. 2. ed. Rio de Janeiro: Ateneu, 2001. ISBN 9788538801023.
- LERCH, Alexander. **An Introduction to audio content analysis. Applications in signal processing and music informatics**. 1st. [S.l.]: IEEE Press, 2012. ISBN 9781118266823.
- LIANG, S. et al. Classification of EEG signals from musicians and non-musicians by neural networks. In: 9TH World Congress on Intelligent Control and Automation. Taipei, Taiwan: [s.n.], jun. 2011. P. 865–869. DOI: 10.1109/WCICA.2011.5970639.
- LOPEZ-CALDERON, Javier; LUCK, Steven J. ERPLAB: an open-source toolbox for the analysis of event-related potentials. **Frontiers in Human Neuroscience**, v. 8, p. 213, 2014. DOI: 10.3389/fnhum.2014.00213.
- LU, Haiping; PLATANIOTIS, Konstantinos N.; VENETSANOPOULOS, Anastasios. **Multilinear Subspace Learning: Dimensionality Reduction of Multidimensional Data**. 1st. [S.l.]: Chapman e Hall/CRC, 2013.
- LUCK, Steven J. **An introduction to the Event-Related Potential Technique**. 2nd. [S.l.]: MIT Press, 2014. ISBN 9780262525855.
- MARCUSE, Lara V.; FIELDS, Madeline C.; YOO, Jiyeoun. **Rowan's Primer of EEG**. 2nd. [S.l.]: Elsevier, 2016. ISBN 9780323353878.
- MARKOVIC, Andjela; KÜHNIS, Jürg; JÄNCKE, Lutz. Task context influences brain activation during music listening. **Frontiers in Human Neuroscience**, v. 11, p. 342, 2017. DOI: 10.3389/fnhum.2017.00342.
- MIKUTTA, C.A. et al. Professional musicians listen differently to music. **Neuroscience**, v. 268, p. 102–111, 2014. ISSN 0306-4522. DOI: 10.1016/j.neuroscience.2014.03.007.
- MITROVIĆ, Dalibor; ZEPPELZAUER, Matthias; BREITENEDER, Christian. Chapter 3 - Features for Content-Based Audio Retrieval. In: **ADVANCES in Computers: Improving the Web**. [S.l.]: Elsevier, 2010. v. 78. (Advances in Computers). P. 71–150. DOI: 10.1016/S0065-2458(10)78003-7.
- MORAES, Dênis de. **O rebelde do traço: a vida de Henfil**. 1st ed. [S.l.]: José Olympio, 2016. ISBN 9788503012737.
- MORENO, Sylvain; BIDELMAN, Gavin M. Examining neural plasticity and cognitive benefit through the unique lens of musical training. **Hearing Research**, v. 308, p. 84–97, 2014. Music: A window into the hearing brain. DOI: 10.1016/j.heares.2013.09.012.
- MÜNTE, Thomas F; ALTENMÜLLER, Eckart; JÄNCKE, Lutz. The musician's brain as a model of neuroplasticity. **Nature reviews Neuroscience**, v. 3, n. 6, p. 473–478, 2002. DOI: 10.1038/nrn843.

- MUSZKAT, Mauro; CORREIA, Cleo M. F.; CAMPOS, Sandra M. Música e Neurociência. **Rev. Neurociências**, v. 8, n. 2, p. 70–75, 2000. DOI: 10.34024/rnc.2000.v8.8947.
- NAGER, Wido et al. The fate of sounds in conductors' brains: an ERP study. **Cognitive Brain Research**, v. 17, n. 1, p. 83–93, 2003. DOI: 10.1016/S0926-6410(03)00083-1.
- NIKJEH, Dee A.; LISTER, Jennifer J.; FRISCH, Stefan A. Hearing of note: An electrophysiologic and psychoacoustic comparison of pitch discrimination between vocal and instrumental musicians. **Psychophysiology**, v. 45, n. 6, p. 994–1007, 2008. DOI: 10.1111/j.1469-8986.2008.00689.x.
- O'BRIEN, Jennifer L.; NIKJEH, Dee A.; LISTER, Jennifer J. Interaction of musicianship and aging: a comparison of cortical auditory evoked potentials. **Behavioural Neurology**, v. 2015, p. 1–12, 2015. DOI: 10.1155/2015/545917.
- OLIVEIRA, Marcelo. **Neurônios**. 2006. Disponível em: <<http://www.infoescola.com/sistema-nervoso/neuronios/>>. Acesso em: 14 set. 2018.
- OPENBCI. **OpenBCI - Open Source Biosensing tools**. 2017. Disponível em: <<http://openbci.com>>. Acesso em: 1 nov. 2017.
- PANTEV, Christo et al. Increased auditory cortical representation in musicians. **Nature**, v. 392, n. 4, p. 811–814, 1998. DOI: 10.1038/33918.
- PANTEV, Christo et al. Timbre-specific enhancement of auditory cortical representations in musicians. **Neuroreports**, v. 12, n. 1, p. 169–174, 2001. DOI: 10.1097/00001756-200101220-00041.
- PASCUAL-LEONE, Alvaro et al. THE PLASTIC HUMAN BRAIN CORTEX. **Annual Review of Neuroscience**, v. 28, n. 1, p. 377–401, 2005. DOI: 10.1146/annurev.neuro.27.070203.144216.
- PATTERSON, Roy D. et al. The Processing of Temporal Pitch and Melody Information in Auditory Cortex. **Neuron**, v. 36, n. 4, p. 767–779, 2002. DOI: 10.1016/S0896-6273(02)01060-7.
- PERETZ, Isabelle; ZATORRE, Robert J. Brain Organization for Music Processing. **Annual Review of Psychology**, v. 56, n. 1, p. 89–114, 2005. DOI: 10.1146/annurev.psych.56.091103.070225.
- POIKONEN, H. et al. Event-related brain responses while listening to entire pieces of music. **Neuroscience**, v. 312, n. 15, p. 58–73, jan. 2016. DOI: 10.1016/j.neuroscience.2015.10.061.
- POIKONEN, Hanna; TOIVAINEN, Petri; TERVANIEMI, Mari. Early auditory processing in musicians and dancers during a contemporary dance piece. **Scientific Reports**, v. 6, set. 2016. DOI: 10.1038/srep33056.
- POLICH, John. Updating P300: An Integrative Theory of P3a and P3b. **Clinical neurophysiology**, v. 118, n. 10, p. 2128–2148, 2009. DOI: 10.1016/j.clinph.2007.04.019.
- RAUSCHECKER, Josef P. Chapter 23 - FUNCTIONAL ORGANIZATION AND PLASTICITY OF AUDITORY CORTEX. In: PERETZ, Isabelle; ZATORRE, Robert J. (Ed.). **The cognitive neuroscience of music**. [S.l.]: Oxford Scholarship, 2003. P. 257–265. DOI: 10.1093/acprof:oso/9780198525202.003.0023.
- RENTFROW, Peter J.; GOSLING, Samuel D. The do re mi's of everyday life: The structure and personality correlates of music preferences. **Journal of Personality and Social Psychology**, v. 84, n. 6, p. 1236–1256, 2003. DOI: 10.1037/0022-3514.84.6.1236.

RIBEIRO, Estela. **Um estudo sobre predição de musicalidade por meio da análise de sinais de EEG**. 2017. Diss. (Mestrado) – Centro Universitário FEI.

RIBEIRO, Estela; THOMAZ, Carlos Eduardo. A Whole Brain EEG Analysis of Musicianship. **Music Perception: An Interdisciplinary Journal**, University of California Press Journals, v. 37, n. 1, p. 42–56, 2019. DOI: 10.1525/mp.2019.37.1.42.

ROCHA, Fábio Theoto et al. Brain Mapping and Interpretation of Reading Processing in Children Using EEG and Multivariate Statistical Analysis. In: 27TH SIBGRAPI, Conference on Graphics, Patterns and Images - IEEE CS Press. Rio de Janeiro, Brazil: [s.n.], ago. 2014. P. 251–258. DOI: 10.1109/SIBGRAPI.2014.19.

ROUSSEEUW, Peter J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. **Journal of Computational and Applied Mathematics**, v. 20, p. 53–65, 1987. DOI: 10.1016/0377-0427(87)90125-7.

SAARI, Pasi et al. Decoding musical training from dynamic processing of musical features in the brain. **Scientific Report**, v. 708, n. 8, p. 1–12, 2018. DOI: 10.1038/s41598-018-19177-5.

SAMSON, SÉVERINE. Neuropsychological Studies of Musical Timbre. **Annals of the New York Academy of Sciences**, v. 999, n. 1, p. 144–151, 2003. DOI: 10.1196/annals.1284.016.

SAMSON, SÉVERINE; EHRLÉ, NATHALIE; BAULAC, MICHEL. Cerebral Substrates for Musical Temporal Processes. **Annals of the New York Academy of Sciences**, v. 930, n. 1, p. 166–178, 2001. DOI: 10.1111/j.1749-6632.2001.tb05732.x.

SANJU, Himanshu Kumar; KUMAR, Prawin. Enhanced auditory evoked potentials in musicians: A review of recent findings. **Journal of Otology**, v. 11, n. 2, p. 63–72, 2016. DOI: 10.1016/j.joto.2016.04.002.

SARKAMO, Teppo. Music for the aging brain: cognitive, emotional, social and neural benefits of musical leisure activities in stroke and dementia. **Dementia**, v. 17, n. 6, p. 670–685, 2017. DOI: 10.1177/1471301217729237.

SATO, Joao Ricardo et al. Hyperplane navigation: A method to set individual scores in fMRI group datasets. **NeuroImage**, v. 42, n. 4, p. 1473–1480, out. 2008. DOI: 10.1016/j.neuroimage.2008.06.024.

SCHLAUG, GOTTFRIED. The Brain of Musicians. **Annals of the New York Academy of Sciences**, v. 930, n. 1, p. 281–299, 2001. DOI: 10.1111/j.1749-6632.2001.tb05739.x.

SCHLAUG, Gottfried. Chapter 3 - Musicians and music making as a model for the study of brain plasticity. In: ALTENMÜLLER, Eckart; FINGER, Stanley; BOLLER, François (Ed.). **Music, Neurology, and Neuroscience: Evolution, the Musical Brain, Medical Conditions, and Therapies**. [S.l.]: Elsevier, 2015. v. 217. (Progress in Brain Research). P. 37–55. DOI: 10.1016/bs.pbr.2014.11.020.

SCHLAUG, Gottfried et al. Effects of Music Training on the Child's Brain and Cognitive Development. **Annals of the New York Academy of Sciences**, v. 1060, n. 1, p. 219–230, 2005. DOI: 10.1196/annals.1360.015.

SCHLAUG, Gottfried et al. Increased corpus callosum size in musicians. **Neuropsychologia**, v. 33, n. 8, p. 1047–1055, 1995. Neuropsychological And Developmental Studies Of The Corpus Callosum. ISSN 0028-3932. DOI: 10.1016/0028-3932(95)00045-5.

SCHNEIDER, Peter et al. Morphology of Heschl's gyrus reflects enhanced activation in the auditory cortex of musicians. **Nature Neuroscience**, v. 5, n. 6, p. 688–694, jun. 2002. DOI: 10.1038/nn871.

- SCHOME, Donald L.; SILVA, Fernando H. Lopes Da. **Niedermeyer's Electroencephalography: Basic Principles, Clinical Applications, and Related Fields**. 6th. [S.l.]: Lippincott Williams e Wilkins, 2010.
- SCHWENKREIS, Peter et al. Assessment of sensorimotor cortical representation asymmetries and motor skills in violin players. **European Journal of Neuroscience**, v. 26, n. 11, p. 3291–3302, 2007. DOI: 10.1111/j.1460-9568.2007.05894.x.
- SEPPÄNEN, M.; BRATTICO, E.; TERVANIEMI, M. Practice strategies of musicians modulate neural processing and the learning of sound-patterns. **Neurobiology of learning and memory**, v. 87, p. 236–247, 2007. DOI: 10.1016/j.nlm.2006.08.011.
- SETHARES, William A. **Tuning, timbre, spectrum, scale**. 2nd ed. Madison: Springer, 2005. ISBN 1852337974.
- SHAHIN, Antoine et al. Enhancement of Neuroplastic P2 and N1c Auditory Evoked Potentials in Musicians. **Journal of Neuroscience**, v. 23, n. 13, p. 5545–5552, 2003. DOI: 10.1523/JNEUROSCI.23-13-05545.2003.
- SLUMING, Vanessa et al. Voxel-Based Morphometry Reveals Increased Gray Matter Density in Broca's Area in Male Symphony Orchestra Musicians. **NeuroImage**, v. 17, n. 3, p. 1613–1622, 2002. DOI: 10.1006/nimg.2002.1288.
- SQUIRE, Larry R. et al. **Fundamental neuroscience**. 3rd ed. [S.l.]: Elsevier, 2008.
- STEFFENER, Jason; STERN, Yaakov. Exploring the neural basis of cognitive reserve in aging. **Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease**, v. 1822, n. 3, p. 467–473, 2012. Imaging Brain Aging and Neurodegenerative Disease. DOI: 10.1016/j.bbadis.2011.09.012.
- STERN, Yaakov. Cognitive Reserve. **Neuropsychologia**, v. 47, n. 10, p. 2015–2028, 2009. DOI: 10.1016/j.neuropsychologia.2009.03.004.
- STURM, Bob L. The State of the Art Ten Years After a State of the Art: Future Research in Music Information Retrieval. **Journal of New Music Research**, Informa UK Limited, v. 43, n. 2, p. 147–172, abr. 2014. DOI: 10.1080/09298215.2014.894533.
- TERVANIEMI, Mari. Musicians - Same or different. **Annals of the New York Academy of Sciences**, v. 1169, p. 151–156, 2009. DOI: 10.1111/j.1749-6632.2009.04591.x.
- TERVANIEMI, Mari et al. Auditory Profiles of Classical, Jazz, and Rock Musicians: Genre-Specific Sensitivity to Musical Sound Features. **Frontiers in Psychology**, v. 6, p. 1900, 2016. DOI: 10.3389/fpsyg.2015.01900.
- TERVANIEMI, Mari et al. Sound processing in amateur musicians and nonmusicians: Event-related potential and behavioral indices. **NeuroReports**, v. 17, n. 11, p. 1225–1228, 2006. DOI: 10.1097/01.wnr.0000230510.55596.8b.
- THOMAZ, Carlos E. et al. Multivariate Statistical Differences of MRI Samples of the Human Brain. **Journal of Mathematical Imaging and Vision**, v. 29, n. 2-3, p. 95–106, 2007. DOI: 10.1007/s10851-007-0033-6.
- TRAINOR, Laurel J.; DESJARDINS, Renée N.; ROCKEL, Conrad. A Comparison of Contour and Interval Processing in Musicians and Nonmusicians Using Event-Related Potentials. **Australian Journal of Psychology**, v. 51, n. 3, p. 147–153, 1999. DOI: 10.1080/00049539908255352.

- TZANETAKIS, G.; COOK, P. Musical genre classification of audio signals. **IEEE Transactions on Speech and Audio Processing**, v. 10, n. 5, p. 293–302, jul. 2002. DOI: 10.1109/TSA.2002.800560.
- UNITED NATIONS. World Population Ageing. **Department of Economic and Social Affairs: Population Division**, 2015.
- UNITED NATIONS. World Population Prospects: The 2017 Revision. **Department of Economic and Social Affairs: Population Division**, 2017.
- VANZELLA, Patricia et al. fNIRS response in professional violinists while playing duets: evidence for distinct leader and follower roles at the brain level. **Frontiers in Psychology**, v. 10, p. 164, 2019. DOI: 10.3389/fpsyg.2019.00164.
- VINCIGUERRA, Claudia. Music intervention efficacy in elderly: a promising non-pharmacological approach to cognitive dysfunctions. **Neurological sciences**, v. 38, n. 6, p. 933–934, 2017. DOI: 10.1007/s10072-017-3011-9.
- VIRTALA, P. et al. Musicianship facilitates the processing of Western music chords - An ERP and behavioral study. **Neuropsychologia**, v. 61, p. 247–258, ago. 2014. DOI: 10.1016/j.neuropsychologia.2014.06.028.
- VUUST, Peter et al. The sound of music: Differentiating musicians using a fast, musical multi-feature mismatch negativity paradigm. **Neuropsychologia**, v. 50, n. 7, p. 1432–1443, jun. 2012. DOI: 10.1016/j.neuropsychologia.2012.02.028.
- WECKER, Jonas Edison; SOARES, Marcelo Marques; NEMOS, Douglas Fabiano Lenz. **Aula de anatomia**. 2001. Disponível em: <<http://www.auladeanatomia.com/novosite/sistemas/sistema-nervoso/>>. Acesso em: 13 set. 2018.
- WOODMAN, Geoffrey F. A Brief Introduction to the Use of Event-Related Potentials (ERPs) in Studies of Perception and Attention. **Attention, Perception and Psychophysics**, v. 72, n. 8, p. 2031–2046, 2010. DOI: 10.3758/BF03196680.
- XAVIER, I. et al. A Photo-Realistic Generator of Most Expressive and Discriminant Changes in 2D Face Images. In: 2015 Sixth International Conference on Emerging Security Technologies (EST). Braunschweig, Germany: [s.n.], set. 2015. P. 80–85. DOI: 10.1109/EST.2015.17.
- YANG, Jinbiao; ZHU, Hao; TIAN, Xing. Group-Level Multivariate Analysis in EasyEEG Toolbox: Examining the Temporal Dynamics Using Topographic Responses. **Frontiers in Neuroscience**, v. 12, p. 468, 2018. DOI: 10.3389/fnins.2018.00468.
- ZATORRE, Robert J.; BELIN, Pascal. Spectral and Temporal Processing in Human Auditory Cortex. **Cerebral Cortex**, v. 11, n. 10, p. 946–953, out. 2001. DOI: 10.1093/cercor/11.10.946.
- ZATORRE, Robert J.; BELIN, Pascal; PENHUNE, Virginia B. Structure and function of auditory cortex: music and speech. **Trends in cognitive science**, v. 6, n. 1, p. 37–46, 2002. DOI: 10.1016/S1364-6613(00)01816-7.
- ZATORRE, Robert J.; CHEN, Joyce L.; PENHUNE, Virginia B. When the brain plays music: auditory–motor interactions in music perception and production. **Nature Reviews Neuroscience**, v. 8, p. 547–558, 2007. DOI: 10.1038/nrn2152.
- ZENDEL, Benjamin Rich; ALAIN, Claude. Musicians experience less age-related decline in central auditory processing. **Psychology and Aging**, v. 27, n. 2, p. 410–417, 2012. DOI: 10.1037/a0024816.

## **APÊNDICE A – DESCRIÇÃO DAS CARACTERÍSTICAS ACÚSTICAS**

Abaixo são descritas as 12 características acústicas utilizadas neste trabalho (LERCH, 2012; KNESS; SCHEDL, 2016; MITROVIĆ; ZEPPELZAUER; BREITENEDER, 2010). Uma descrição detalhada da extração de cada uma delas pode ser encontrada no manual da MIRtoolbox (LARTILLOT, 2014), assim como nos livros Kness e Schedl (2016) e Lerch (2012).

### *Root Mean Square (RMS)*

Relacionado à Intensidade do sinal de áudio, o RMS é a medida da energia do sinal, computada por meio da raiz quadrada média do quadrado da amplitude do sinal. O valor do RMS pode ser extraído a partir do sinal de áudio  $x(n)$ , de acordo com a Equação (16), com uma janela de comprimento N, tal que

$$RMS = \sqrt{\frac{1}{N} \sum_{n=0}^{\frac{N}{2}} x(n)^2}. \quad (16)$$

### *Zero Crossing Rate (ZCR)*

Relacionado a Tonalidade, o ZCR é a medida do número de vezes que o sinal cruza o eixo do tempo. O valor de ZCR pode ser extraído a partir do sinal de áudio  $x(n)$ , de acordo com a Equação

$$ZCR = \frac{1}{2N} \sum_{n=0}^{\frac{N}{2}} |\text{sign}(x(n)) - \text{sign}(x(n-1))|, \quad (17)$$

onde a função  $\text{sign}$  é definida como

$$\text{sign}(x(n)) = \begin{cases} 1, & \text{se } x(n) \geq 0 \\ -1, & \text{se } x(n) < 0 \end{cases}, \quad (18)$$

onde  $x(n-1) = 0$  na inicialização ou quando tal valor não existe.

### *Spectral Rolloff*

Relacionado a Tonalidade, é a frequência ( $M$ ) abaixo da qual 85% ( $\beta$ ) da energia total está contida no sinal. Assim, temos

$$SRO = M, \quad (19)$$

tal que:

$$\sum_{k=0}^{\frac{M}{2}} |X(k)| = \beta \sum_{k=0}^{\frac{N}{2}} |X(k)|, \quad (20)$$

onde  $X(k)$  é a FFT do sinal de áudio  $x(n)$ .

### *Spectral Roughness*

Relacionado a Dissonância, é a estimativa da dissonância sensorial. Ou seja, quando dois tons soam em frequências muito próximas. Quando a diferença entre esses dois tons aumenta, a sensação de aspereza (dissonância) do som diminui. Assim, o spectral roughness estima a dissonância dependendo da razão entre a frequência de pares de ondas senoidais, sendo que quando a diferença de frequência entre dois tons é de aproximadamente 32Hz, ocorre a dissonância (SETHARES, 2005).

### *Brightness*

Relacionado ao Brilho, é a medida da quantidade de energia acima de 1500 Hz. Quanto maior o seu valor, maior a energia do sinal e maior a intensidade do brilho na janela calculada.

### *Spectral Entropy*

Relacionado a Tonalidade, é a medida relativa a Entropia de Shannon do sinal, indicando se o espectro contém picos predominantes ou não. Se a curva é achatada, correspondendo a uma situação de máxima incerteza, então a entropia é máxima. Em contrapartida, um tom puro possui mínima entropia. A spectral entropy pode ser calculada como

$$SE = \frac{\sum_{k=0}^{\frac{N}{2}} X(k) \cdot \log X(k)}{\log \frac{N}{2}}, \quad (21)$$

onde  $X(k)$  é a FFT do sinal de áudio  $x(n)$ .

### *Spectral Flatness*

Relacionado a Tonalidade, é a medida da uniformidade do espectro, definido como a razão entre a média geométrica e a média aritmética do sinal, também conhecida como Entropia de Wiener. Valores altos indicam achatamento do espectro, indicando ruído, enquanto valores baixos indicam não achatamento do espectro, indicando ondas senoidais. O spectral flatness é calculado como

$$SF = \frac{\sqrt[N]{\prod_{k=0}^{\frac{N}{2}} |X(k)|}}{\frac{1}{N} \sum_{k=0}^{\frac{N}{2}} |X(k)|}, \quad (22)$$

onde  $X(k)$  é a FFT do sinal de áudio  $x(n)$ .

### *Spectral Skewness*

Relacionado a Forma Espectral, é o terceiro momento central da distribuição do espectro do sinal de áudio, relacionado com a assimetria da distribuição. O spectral skewness é calculado como

$$Skw = \frac{\frac{1}{N} \sum_{k=0}^{\frac{N}{2}} (X(k) - \bar{X})^3}{\left(\sqrt{\frac{1}{N} \sum_{k=0}^{\frac{N}{2}} (X(k) - \bar{X})^2}\right)^3}, \quad (23)$$

onde  $\bar{X}$  é definido como:

$$\bar{X} = \frac{1}{N} \sum_{k=0}^{\frac{N}{2}} X(k), \quad (24)$$

onde  $X(k)$  é a FFT do sinal de áudio  $x(n)$ .

### *Spectral Kurtosis*

Relacionado a Forma Espectral, é o quarto momento central da distribuição do espectro, indica o achatamento do espectro e mudanças súbitas podem indicar transientes no áudio. O spectral kurtosis é calculado como

$$Kur = \frac{\frac{1}{N} \sum_{k=0}^{\frac{N}{2}} (X(k) - \bar{X})^4}{\left(\frac{1}{N} \sum_{k=0}^{\frac{N}{2}} (X(k) - \bar{X})^2\right)^2} - 3, \quad (25)$$

onde  $\bar{X}$  é definido pela Equação (24) e  $X(k)$  é a FFT do sinal de áudio  $x(n)$ .

### *Spectral Centroid*

Relacionado a Forma Espectral e ao Brilho, é o primeiro momento central da distribuição do espectro, é o centro geométrico da distribuição, relacionado com as propriedades de brilho e nitidez do áudio, sendo calculado como

$$SC = \frac{\sum_{k=0}^{\frac{N}{2}} kX(k)}{\sum_{k=0}^{N-1} X(k)}, \quad (26)$$

onde  $X(k)$  é a FFT do sinal de áudio  $x(n)$ .

### *Spectral Spread*

Relacionado a Tonalidade e a Forma Espectral, é o segundo momento central da distribuição do espectro. Sons tonais apresentam baixos valores, enquanto sons ruidosos apresentam altos valores. O spectral spread pode ser calculado como

$$SS = \sqrt{\frac{\sum_{k=0}^{\frac{N}{2}} |k - SC| X(k)}{\sum_{k=0}^{\frac{N}{2}} X(k)}}, \quad (27)$$

onde  $X(k)$  é a FFT do sinal de áudio  $x(n)$  e  $SC$  é o *Spectral Centroid* calculado pela equação anterior (26).

### *Spectral Flux*

É a medida das mudanças temporais no espectro entre janelas sucessivas. Sinais com baixa variação (ou quase constantes) como sons ruidosos apresentam baixos valores, enquanto mudanças abruptas no espectro (e.g. note onset) apresentam altos valores. O spectral flux pode ser considerado uma aproximação rudimentar da sensação de dissonância (LERCH, 2012) e é calculado como

$$SFlux = \sum_{k=0}^{\frac{N}{2}} (|X(w, k) - X(w - 1, k)|)^2, \quad (28)$$

onde  $X(w, k)$  é a FFT da janela  $w$  e  $X(w - 1, k)$  é a FFT da janela  $w - 1$  do sinal de áudio  $x(n)$ .

## **APÊNDICE B – TABELAS**

Tabela 5 – Carga fatorial das variáveis para o gênero *Blues*

	<b>F1</b>	<b>F2</b>
<b>RMS</b>	0.052238	0.92118
<b>ZCR</b>	0.75878	0.21914
<b>S. Rolloff</b>	0.95237	0.099256
<b>S. Roughness</b>	0.13857	0.84107
<b>Brightness</b>	0.89518	0.24394
<b>S. Entropy</b>	0.90126	0.23125
<b>S. Flatness</b>	0.89119	0.13419
<b>S. Skewness</b>	-0.95054	-0.12743
<b>S. Kurtosis</b>	-0.84646	-0.087734
<b>S. Centroid</b>	0.9514	0.16199
<b>S. Spread</b>	0.86609	0.0083685
<b>S. Flux</b>	0.20969	0.83372

Tabela 6 – Carga fatorial das variáveis para o gênero *Classical*

	<b>F1</b>	<b>F2</b>
<b>RMS</b>	0.077455	0.95006
<b>ZCR</b>	0.76237	0.15001
<b>S. Rolloff</b>	0.95725	0.092974
<b>S. Roughness</b>	0.089535	0.91179
<b>Brightness</b>	0.90303	0.14682
<b>S. Entropy</b>	0.75895	0.21043
<b>S. Flatness</b>	0.76373	0.19848
<b>S. Skewness</b>	-0.90918	-0.051797
<b>S. Kurtosis</b>	-0.85877	-0.034324
<b>S. Centroid</b>	0.94494	0.13928
<b>S. Spread</b>	0.84246	0.050944
<b>S. Flux</b>	0.21101	0.90841

Tabela 7 – Carga fatorial das variáveis para o gênero *Country*

	<b>F1</b>	<b>F2</b>
<b>RMS</b>	0.075192	0.92294
<b>ZCR</b>	0.77597	0.17351
<b>S. Rolloff</b>	0.95433	0.12849
<b>S. Roughness</b>	0.15714	0.84947
<b>Brightness</b>	0.90655	0.232
<b>S. Entropy</b>	0.91409	0.23488
<b>S. Flatness</b>	0.90622	0.098538
<b>S. Skewness</b>	-0.91516	-0.19392
<b>S. Kurtosis</b>	-0.79244	-0.16045
<b>S. Centroid</b>	0.9363	0.15261
<b>S. Spread</b>	0.91034	0.048609
<b>S. Flux</b>	0.21254	0.81321

Tabela 8 – Carga fatorial das variáveis para o gênero *Disco*

	<b>F1</b>	<b>F2</b>
<b>RMS</b>	-0.012566	0.94376
<b>ZCR</b>	0.77733	-0.055013
<b>S. Rolloff</b>	0.93706	0.093081
<b>S. Roughness</b>	0.14948	0.78253
<b>Brightness</b>	0.87679	0.057574
<b>S. Entropy</b>	0.89571	0.087153
<b>S. Flatness</b>	0.83377	0.22545
<b>S. Skewness</b>	-0.93559	-0.039577
<b>S. Kurtosis</b>	-0.83469	-0.075931
<b>S. Centroid</b>	0.86302	0.04774
<b>S. Spread</b>	0.80791	0.15616
<b>S. Flux</b>	0.075417	0.84926

Tabela 9 – Carga fatorial das variáveis para o gênero *Hiphop*

	<b>F1</b>	<b>F2</b>
<b>RMS</b>	-0.15268	0.91684
<b>ZCR</b>	0.72636	-0.22551
<b>S. Rolloff</b>	0.93996	0.083204
<b>S. Roughness</b>	0.089053	0.7544
<b>Brightness</b>	0.88697	-0.16079
<b>S. Entropy</b>	0.83879	-0.054911
<b>S. Flatness</b>	0.8678	0.14165
<b>S. Skewness</b>	-0.91759	0.063853
<b>S. Kurtosis</b>	-0.78527	-0.017878
<b>S. Centroid</b>	0.84228	-0.015002
<b>S. Spread</b>	0.73837	0.27269
<b>S. Flux</b>	0.015272	0.8225

Tabela 10 – Carga fatorial das variáveis para o gênero *Jazz*

	<b>F1</b>	<b>F2</b>
<b>RMS</b>	-0.038074	0.90957
<b>ZCR</b>	0.74253	0.040683
<b>S. Rolloff</b>	0.96276	0.026392
<b>S. Roughness</b>	0.088299	0.83586
<b>Brightness</b>	0.91758	0.1171
<b>S. Entropy</b>	0.89644	0.10418
<b>S. Flatness</b>	0.88741	0.073766
<b>S. Skewness</b>	-0.92196	-0.078989
<b>S. Kurtosis</b>	-0.79056	-0.081972
<b>S. Centroid</b>	0.93942	0.053813
<b>S. Spread</b>	0.89798	-0.0019386
<b>S. Flux</b>	0.12875	0.82919

Tabela 11 – Carga fatorial das variáveis para o gênero *Metal*

	<b>F1</b>	<b>F2</b>
<b>RMS</b>	-0.10997	0.95477
<b>ZCR</b>	0.80957	-0.10429
<b>S. Rolloff</b>	0.92652	0.073899
<b>S. Roughness</b>	0.17271	0.86232
<b>Brightness</b>	0.85427	-0.079634
<b>S. Entropy</b>	0.89787	0.025767
<b>S. Flatness</b>	0.84522	0.13731
<b>S. Skewness</b>	-0.91869	-0.022242
<b>S. Kurtosis</b>	-0.82402	-0.077405
<b>S. Centroid</b>	0.94169	-0.013
<b>S. Spread</b>	0.76666	0.13317
<b>S. Flux</b>	0.025258	0.89358

Tabela 12 – Carga fatorial das variáveis para o gênero *Pop*

	<b>F1</b>	<b>F2</b>
<b>RMS</b>	-0.058072	0.92716
<b>ZCR</b>	0.75062	-0.063054
<b>S. Rolloff</b>	0.92898	0.11539
<b>S. Roughness</b>	0.17564	0.68866
<b>Brightness</b>	0.90617	0.033199
<b>S. Entropy</b>	0.89238	0.059778
<b>S. Flatness</b>	0.90468	0.14077
<b>S. Skewness</b>	-0.96468	-0.10537
<b>S. Kurtosis</b>	-0.82333	-0.14892
<b>S. Centroid</b>	0.6887	0.033875
<b>S. Spread</b>	0.78773	0.17335
<b>S. Flux</b>	0.089211	0.86219

Tabela 13 – Carga fatorial das variáveis para o gênero *Reggae*

	<b>F1</b>	<b>F2</b>
<b>RMS</b>	-0.039853	0.87095
<b>ZCR</b>	0.76091	0.081398
<b>S. Rolloff</b>	0.93981	0.14958
<b>S. Roughness</b>	0.079608	0.67735
<b>Brightness</b>	0.91624	0.089407
<b>S. Entropy</b>	0.91157	0.0748
<b>S. Flatness</b>	0.87178	0.25901
<b>S. Skewness</b>	-0.94514	-0.065043
<b>S. Kurtosis</b>	-0.80773	-0.040024
<b>S. Centroid</b>	0.83934	-0.12851
<b>S. Spread</b>	0.81709	0.17652
<b>S. Flux</b>	0.18897	0.84073

Tabela 14 – Carga fatorial das variáveis para o gênero *Rock*

	<b>F1</b>	<b>F2</b>
<b>RMS</b>	-0.063161	0.93516
<b>ZCR</b>	0.80539	0.078288
<b>S. Rolloff</b>	0.94835	0.0039254
<b>S. Roughness</b>	0.14175	0.86513
<b>Brightness</b>	0.8969	0.1116
<b>S. Entropy</b>	0.90463	0.16848
<b>S. Flatness</b>	0.84762	0.11522
<b>S. Skewness</b>	-0.93334	-0.044169
<b>S. Kurtosis</b>	-0.81792	-0.054899
<b>S. Centroid</b>	0.91284	0.0125
<b>S. Spread</b>	0.82169	-0.026492
<b>S. Flux</b>	0.10044	0.86171