



## CODIFICAÇÃO E DECODIFICAÇÃO DA RESPOSTA CEREBRAL À MÚSICA

Sebastian Höfle

Tese de Doutorado apresentada ao Programa de Pós-graduação em Engenharia Biomédica, COPPE, Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do grau de Doutor em Engenharia Biomédica.

Orientadores: Maurício Cagy

Jorge Neval Moll Neto

Rio de Janeiro

Agosto de 2017

# CODIFICAÇÃO E DECODIFICAÇÃO DA RESPOSTA CEREBRAL À MÚSICA

Sebastian Höfle

TESE SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA (COPPE) DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE DOUTOR EM CIÊNCIAS EM ENGENHARIA BIOMÉDICA.

Examinada por:

---

Prof. Maurício Cagy, D.Sc.

---

Dr. Jorge Neval Moll Neto, D.Sc.

---

Prof. Carlos Julio Tierra Criollo, D.Sc.

---

Prof. Tiago Arruda Sanchez, D.Sc.

---

Prof. Leticia de Oliveira, D.Sc.

RIO DE JANEIRO, RJ BRASIL

AGOSTO DE 2017

Höfle, Sebastian.

Codificação e decodificação da resposta cerebral à música/ Sebastian Hoefle. – Rio de Janeiro: UFRJ/COPPE, 2017.

XIII, 87 f.: il.; 29,7 cm.

Orientadores: Maurício Cagy

Jorge Neval Moll Neto.

Tese (doutorado) – UFRJ/ COPPE/ Engenharia Biomédica, 2017.

Referências Bibliográficas: p. 75-80.

1. RMf. 2. Regressão linear múltipla. 3. Música. I. Cagy, Maurício et al.. II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia Biomédica. III. Título.

## Agradecimentos

Quero agradecer ao meu orientador Prof. Maurício Cagy por todo apoio durante o curso de doutorado e o desenvolvimento dessa tese. Agradeço por ter me orientado com equilíbrio, permitindo a elaboração do meu trabalho com uma independência necessária e ao mesmo tempo apoiando e direcionando quando foi preciso. Tenho em grande estima a sua disponibilidade, boa vontade e seus ensinamentos durante esse tempo todo.

Quero agradecer ao meu orientador Jorge Moll por ter me dado a oportunidade de desenvolver esse trabalho. Sem seu estímulo, apoio e sua contribuição esse trabalho não teria sido realizado. Seu conhecimento e sua expertise foram fundamentais e seu entusiasmo e suas inspirações o motor principal desse trabalho. Foi uma grande honra ter você como orientador e uma oportunidade única de fazer parte do seu grupo de pesquisa.

Agradeço a Annerose Engel que mesmo não sendo uma orientadora oficial, teve um papel fundamental não só para minha evolução acadêmica como também para esse trabalho específico. Com a sua experiência científica foi possível de elaborar as bases desse trabalho. Agradeço pela confiança, motivação e pelo apoio durante esse tempo. Sua postura científica e pessoal foi âncora e referência!

Da mesma forma quero agradecer ao Rodrigo Basilio, cuja genialidade na programação e expertise matemática contribuíram crucialmente a esse trabalho. Agradeço pela disposição para ajudar e pelas discussões que acompanharam e trilharam o caminho.

Quero agradecer a Vinoa Alluri e ao Petri Toiviainen por compartilhar a expertise na área de música e pelas críticas construtivas que beneficiaram esse trabalho.

Quero agradecer aos meus colegas do dia-dia pelo companheirismo e pela confiança: ao meu chefe Ivanei Bramati pela compreensão e apoio, à Debora Lima e equipe dela pelo apoio em diversas questões, ao meu colega de turma Bruno Melo pela referência no pragmatismo, a Marina pelo interesse dela na música e nesse trabalho, a Fernanda Meireles pelos ensinamentos físicos e o Tiago Bortolini pelas discussões neuro-científicas.

Agradeço aos participantes do experimento pela dedicação do seu tempo.

Agradeço ao Instituto de Pesquisa e Ensino IDOR pelo apoio financeiro a mim e ao meu programa, sem o qual este trabalho não poderia ter sido realizado.

Por fim, agradeço à minha família pelo apoio e pela compreensão da minha ausência durante o percurso do doutorado.

Resumo da Tese apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Doutor em Ciências (D.Sc.)

## CODIFICAÇÃO E DECODIFICAÇÃO DA RESPOSTA CEREBRAL À MÚSICA

Sebastian Höfle

Agosto/2017

Orientadores: Maurício Cagy

Jorge Neval Moll Neto

Programa: Engenharia Biomédica

Atividades cerebrais em resposta à música têm sido investigadas de diferentes maneiras nos últimos anos. Porém, o mapeamento contínuo da atividade cerebral ao ouvir música, em combinação com a modelagem de características musicais, é uma abordagem recente. Os chamados “modelos de codificação” permitem capturar os efeitos de múltiplas variáveis do estímulo na resposta cerebral. Ainda, estes modelos podem ser utilizados para decodificar ou identificar estímulos a partir da atividade cerebral. Este trabalho utiliza essas novas abordagens de codificação e decodificação da atividade cerebral em resposta à música. Primeiro, características musicais extraídas do sinal de áudio são mapeadas para a atividade cerebral, registrada por ressonância magnética funcional. A partir desse mapeamento, as representações espaciais e multivariadas são decodificadas para identificar qual música foi escutada. Numa investigação sistemática, os parâmetros internos da modelagem que levam ao desempenho ótimo do modelo são revelados e discutidos. Entre os parâmetros mais importantes são a escolha ideal da dimensão espacial e a consideração da entropia das músicas, resultando em acurácias máximas de até 95%.

Abstract of Thesis presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Doctor of Science (D.Sc.)

## ENCODING AND DECODING BRAIN ACTIVITY IN RESPONSE TO MUSIC

Sebastian Höfle

August/2017

Advisors: Maurício Cagy

Jorge Neval Moll Neto

Department: Biomedical Engineering

The neural correlates of listening to music have been investigated in several ways. However, mapping ongoing brain activity during naturalistic music listening combined with detailed models of musical features is an emerging approach. The so-called “encoding models” allow capturing the effects of multiple stimulus variables on brain responses that can be used subsequently to decode or identify stimuli from brain activity. This work applies methods for encoding and decoding brain activity in response to naturalistic music listening. First, musical features are extracted from the audio signal and mapped to brain activity, measured by functional magnetic resonance imaging. Building on this mapping, multivariate spatial representations are decoded in order to identify a specific music which is listened as the activity is measured. Further, a systematic investigation reveals internal parameters that maximize model performance. Among the most important parameters are the ideal point in the spatial dimension and the consideration of entropy of the music pieces, resulting in maximum accuracies of up to 95%.

## Sumário

Lista de figuras .....	xi
Lista de tabelas .....	xiii
1 Introdução .....	1
1.1 A música e o cérebro .....	1
1.2 Mapeando a dinâmica da atividade cerebral.....	2
1.3 Motivação .....	5
1.4 Organização do texto .....	6
2 Fundamentação teórica .....	8
2.1 Teorias de emoções (musicais).....	8
2.1.1 Modelos dimensionais e categorias de emoções .....	10
2.1.2 Emoções e alterações fisiológicas .....	11
2.2 Respostas cerebrais a música.....	12
2.3 Ressonância magnética.....	15
2.3.1 História de RM .....	15
2.3.2 Sistema de RM .....	16
2.3.3 Formação de imagem.....	17
2.3.4 Sinal BOLD .....	17
2.4 Codificação e decodificação respostas cerebrais.....	18
2.4.1 Domínio visual .....	18
2.4.2 Domínio auditivo.....	21
2.4.3 Domínio da música.....	23
3 Métodos .....	28
3.1 Participantes.....	28
3.2 Estímulo.....	28



3.3	Protocolo experimental.....	30
3.3.1	Parâmetros de aquisição de imagens .....	32
3.4	Pré-processamento RMf .....	32
3.4.1	Inspeção visual do dado bruto .....	33
3.4.2	Realinhamento .....	34
3.4.3	Remoção de tendências/frequências baixas.....	34
3.4.4	Correção por efeitos fisiológicos .....	34
3.4.5	Correção do sinal global (cérebro inteiro).....	36
3.4.6	Suavização espacial .....	36
3.4.7	Suavização temporal.....	36
3.4.8	Normalização linear para espaço MNI .....	36
3.5	Codificação de características musicais .....	37
3.5.1	Extração de características acústicas das músicas.....	37
3.5.2	Regressão linear múltipla .....	39
3.5.3	Validação cruzada.....	40
3.5.4	Metodologia “Maestro qual é a música?”.....	40
4	Resultados.....	45
4.1	Escalas comportamentais.....	45
4.2	Respostas fisiológicas durante RMf .....	47
4.3	Comparação de modelos.....	48
4.3.1	Comparação: correlações BOLD-BOLD e codificação .....	49
4.3.2	Comparação de modelos de codificação .....	51
4.4	Aplicação: “Maestro, qual é a música?” .....	53
5	Discussão.....	66
6	Conclusão .....	74
7	Referências bibliográficas .....	75

		x
8	Apêndice.....	81
8.1	Correção do sinal global.....	81
8.2	Telas da pontuação .....	83
8.3	Questionário pós-RM .....	86

## Lista de figuras

Figura 1-1: Ilustração sobre codificação e decodificação. ....	5
Figura 2-1: Primeira imagem obtida com RM. ....	16
Figura 3-1: Protocolo experimental. ....	31
Figura 3-2: Visualização do dado bruto no espaço original da aquisição: .....	33
Figura 3-3: Modelo de codificação e decodificação.....	42
Figura 3-4: Identificação de músicas.....	43
Figura 4-1: Box-plots mostrando pontuações individuais antes da RM.....	45
Figura 4-2: Box-plot mostrando a intensidade da emoção sentida.....	46
Figura 4-3: Pontuações da escala GEMS-9 .....	47
Figura 4-4: Respostas fisiológicas .....	48
Figura 4-5: Comparação da correlação de predição da codificação .....	50
Figura 4-6: Comparação do modelo original.....	52
Figura 4-7: Acurácias de identificação .....	54
Figura 4-8: Acurácia média de seis sujeitos .....	55
Figura 4-9: Voxels utilizados durante a identificação .....	56
Figura 4-10: Distribuição cortical das características acústicas .....	58
Figura 4-11: Distribuição cortical individual dos componentes.....	59
Figura 4-12: Relação da acurácia de identificação com o número de variáveis. 61	
Figura 4-13: Robustez do modelo na identificação .....	62
Figura 4-14: Comparação de identificação.....	63
Figura 4-15: Similaridade de características acústicas .....	64
Figura 4-16: Entropia como fator importante na acurácia:.....	65
Figura 5-1: A distribuição das correlações de predição .....	70
Figura 5-2: Correlações de predição para características não-correspondentes: 71	

Figura 8-1: Distribuição das correlações do sinal global .....	81
Figura 8-2: Distribuição das correlações depois da correção do sinal global ....	82
Figura 8-3: Diferenças entre a tarefa de sentir e não sentir .....	82
Figura 8-4: Efeito da remoção do sinal global na codificação. ....	83
Figura 8-5: Tela de pontuação: Intensidade da emoção .....	83
Figura 8-6: Tela de pontuação: Tarefa técnica para não sentir a emoção .....	84
Figura 8-7: Escala de GEMS de 9 itens antes de pontuar .....	84
Figura 8-8: Escala de GEMS com a pontuação realizada .....	85
Figura 8-9: Questionário preenchido depois de cada sessão de RM. ....	87

## Lista de tabelas

Tabela 2-1 – Modelo da emoção em componentes de processo.....	9
Tabela 2-2 – Relação de número de variáveis, pontos e modelos em publicações de codificação .....	26
Tabela 3-1 – Descrição das características musicais .....	38

# 1 Introdução

## 1.1 A música e o cérebro

Nos últimos anos, emergiram muitos estudos que investigaram a resposta cerebral à música. Os estudos abordaram assuntos diversos desde a percepção básica de características musicais (KOELSCH, 2011), resposta emocional (KOELSCH, 2010, 2014), até efeitos terapêuticos da música (KOELSCH, 2009). Os correlatos neurofisiológicos da percepção musical são resumidos no artigo de (KOELSCH, 2011). Neste trabalho, o autor descreve detalhadamente o fluxo neural da percepção de ritmo, melodia e harmonia que passa pelo tronco cerebral, tálamo e áreas corticais de Brodmann 6, 22, 41, 42, 44, 45, 52. O advento da técnica da ressonância magnética funcional permitiu estudar as respostas em áreas subcorticais. Áreas como amígdala, hipocampo e o núcleo acumbente no estriado ventral são, principalmente, envolvidos nas respostas emocionais à música (KOELSCH, 2014). Vários estudos analisaram diferentes aspectos emocionais como, por exemplo, o prazer da música (SALIMPOOR; ZATORRE, 2013), a valência (KOELSCH et al., 2013; MITTERSCHIFFTHALER et al., 2007; TROST et al., 2011; VIINIKAINEN; KÄTSYRI; SAMS, 2012), a influência da personalidade (PARK et al., 2013) e de doenças psiquiátricas como p.ex. depressão (AUST et al., 2013). Outros estudos focaram mais na relação de modelos emocionais e a música (LEHNE; ROHRMEIER; KOELSCH, 2013; TROST et al., 2011).

Mais recentemente, aumentaram os estudos que modelaram a atividade cerebral a partir de uma descrição quantitativa dos estímulos sonoros. Conhecendo as características de tempo-frequência do estímulo, um estudo (SANTORO et al., 2014) conseguiu identificar sons dentro de um conjunto de 24 sons (que não foram utilizados durante a construção do modelo) a partir da atividade cerebral. De forma similar, outros estudos (ALLURI et al., 2012, 2013) utilizaram uma descrição de estímulos musicais num nível mais alto. Características rítmicas, timbrais e tonais de algumas músicas foram mapeadas na atividade cerebral. A partir desse mapeamento, os autores conseguiram prever a resposta cerebral de outras músicas.

## 1.2 Mapeando a dinâmica da atividade cerebral

A atividade cerebral pode ser registrada em tempo real com diversas técnicas, como p.ex. eletroencefalografia, espectroscopia no infravermelho próximo ou ressonância magnética. Neste trabalho, a técnica de ressonância magnética funcional (RMf) foi utilizada para registrar a atividade cerebral. Embora tenha algumas desvantagens (custo, restrição de movimento/posição, barulho, resolução temporal), ela possui uma resolução espacial de poucos milímetros com a tendência de diminuir ainda mais em aparelhos futuros.

O princípio da ressonância magnética funcional se baseia em características magnéticas que dependem do nível de oxigênio no sangue (inglês: *Blood Oxygen Level Dependent* - BOLD). Quanto mais oxigênio no sangue, menor a susceptibilidade magnética devido à característica química da hemoglobina oxigenada, que é diamagnética. Por outro lado, hemoglobina desoxigenada possui um momento magnético e, por isso, sangue com menos oxigênio resulta numa magnetização maior. Essa magnetização maior diminui a homogeneidade do campo magnético, que resulta num decaimento mais rápido da relaxação transversal, que acontece depois da estimulação dos núcleos pela radiofrequência. Essa relaxação transversal é afetada pela homogeneidade do campo magnético, e é observada no sinal chamado  $T_2^*$  e interpretada como sinal BOLD. Resumindo, o sinal BOLD é menor para hemoglobina desoxigenada devido ao decaimento mais rápido da relaxação  $T_2^*$  e maior para hemoglobina oxigenada. Pelo fato de que a oferta de sangue oxigenado numa área ativa no cérebro é maior do que a mesma área consegue consumir, existe um excedente de oxigênio que, finalmente, resulta num sinal BOLD maior em áreas ativas do cérebro (HUETTEL; SONG; MCCARTHY, 2004).

O sinal BOLD pode ser analisado de diferentes maneiras. Enquanto a maioria das análises de RMf, até hoje, utiliza um modelo linear genérico, contrastando condições bem controladas dentro de um experimento para inferir sobre a atividade cerebral, novas abordagens estão surgindo que tentam prever e decodificar a resposta cerebral a partir de uma descrição mais completa possível de estímulos (NASELARIS et al., 2011). Essas abordagens, muitas vezes, utilizam técnicas de aprendizagem de

máquina para inferir sobre diferentes estados cerebrais (VAROQUAUX; THIRION, 2014). Um conceito central da técnica de aprendizagem de máquina é a validação cruzada. Nessa validação, os dados de treino são separados dos dados de teste. Com base nos dados de treino, o modelo é estimado. De forma mais geral, o modelo relaciona algum *input* de dado com um dado observado. Pode-se pensar no *input* como variáveis independentes, por exemplo, uma descrição dos estímulos por múltiplas variáveis. O dado observado pode ser o sinal BOLD em determinada região ou *voxel*. Em princípio, não há restrições de que forma o modelo relaciona as variáveis independentes com as variáveis dependentes. Porém, a forma mais simples, mais comum e de fácil interpretação é um modelo de regressão linear múltipla. Neste caso, a variável dependente é explicada por uma combinação linear de pesos das variáveis independentes. Uma vez estimada essa combinação nos dados de treino, ela é utilizada para prever a variável dependente nos dados de teste. Dessa forma, o modelo é validado em dados novos, não utilizados durante o treino, evitando problemas de *overfitting* e comprovando (ou não) a validade do modelo em determinada região.

Ainda não existem muitas ferramentas que implementam análises desse tipo. Apenas algumas bibliotecas implementadas em Matlab<sup>®</sup> ou Python ajudam no acesso às imagens de RMf e disponibilizam alguns métodos padrão da área de aprendizagem de máquina (PEDREGOSA et al., 2011).

De forma geral, podem-se categorizar as novas abordagens como:

- a) Codificação (inglês: *Encoding*): é o mapeamento do estímulo externo na atividade cerebral. Geralmente, a codificação é realizada separadamente por *voxel*, isto é, univariado, utilizando a regressão linear múltipla do espaço das características dos estímulos para o sinal BOLD. Às vezes, se utilizam estimadores mais sofisticados como a regressão *ridge* ou *lasso*, por exemplo, no caso em que o número de variáveis é bem grande e da mesma ordem que os números de exemplos disponíveis para o treino do modelo (KAY et al., 2008). Diferentes descrições do mesmo estímulo podem ser utilizadas e comparadas. Pode-se descrever uma imagem ou fotografia em termos de características visuais, por exemplo com *wavelets* de Gabor com diferentes orientações e resoluções espaciais (ÇUKUR et al., 2013; KAY et al., 2008). Por outro lado, uma fotografia pode ser descrita também em um espaço semântico, pelos objetos que estão presentes nela (ÇUKUR et al.,



2013; STANSBURY; NASELARIS; GALLANT, 2013). A comparação de modelos alternativos é interessante por responder quais regiões cerebrais são melhor explicadas por quais modelos. Quanto melhor um modelo prediz a atividade cerebral em determinada região, mais evidente fica a representação desse modelo nesta região.

- b) Decodificação (inglês: *Decoding*): é a inferência sobre o estímulo externo a partir da atividade cerebral. Muitas vezes, essa inferência é realizada de forma multivariada, ou seja, utilizando um conjunto de *voxels* dentro de uma região de interesse (CHIKAZOE et al., 2014; NASELARIS et al., 2011; STANSBURY; NASELARIS; GALLANT, 2013). Essa abordagem é interessante pelas possíveis aplicações no mundo real como, por exemplo, para comandar uma interface cérebro-máquina.
- c) Identificação: é um caso específico da decodificação onde um estímulo dentro de um conjunto finito é identificado.
- d) Reconstrução: pode ser vista como a decodificação mais sofisticada ou a decodificação completa. A partir da atividade cerebral, o estímulo externo é reconstruído no seu formato original. A reconstrução pode ser realizada de diversas maneiras e depende do modelo utilizado. Uma forma é a utilização de um prior composto por um banco de dados grande de estímulos. Nesse caso, a reconstrução combina os estímulos mais prováveis de terem produzidos determinada atividade nos *voxels*. Formalmente, esse processo pode ser descrito com o teorema de Bayes (NASELARIS et al., 2009). Outra forma é a reconstrução direta do estímulo pelo modelo. Por exemplo, após estimar parâmetros que relacionam a atividade cerebral com espectrogramas dos sons durante o treino, esses podem ser utilizados para reconstruir um som, transformando um espectrograma estimado em um sinal de áudio (SANTORO et al., 2017).

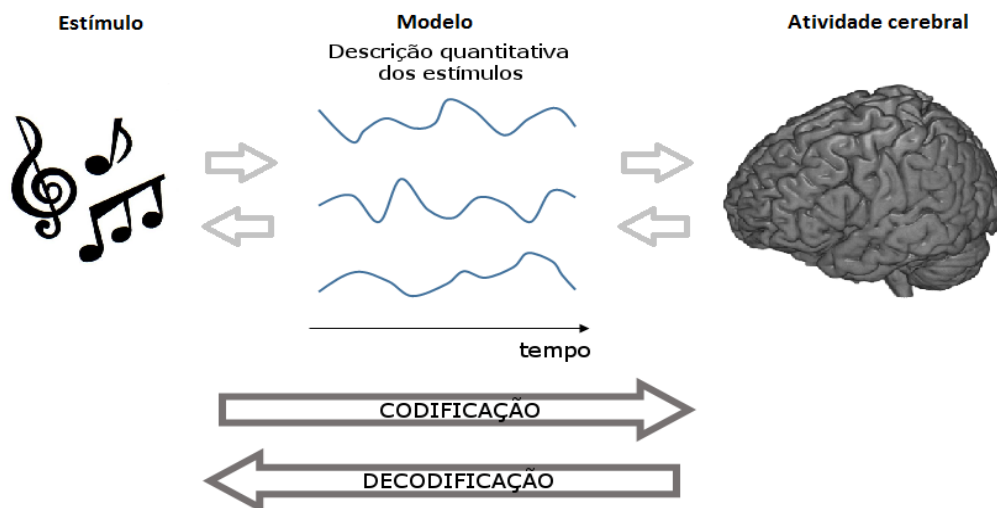


Figura 1-1: Ilustração sobre codificação e decodificação.

Enquanto a codificação é o mapeamento do estímulo para a atividade cerebral, a decodificação vai na direção contrária: a partir da atividade cerebral, é deduzido o estímulo, ou seja, decodificada a atividade.

### 1.3 Motivação

A modelagem de codificação permite prever a atividade cerebral a partir de uma descrição do estímulo externo. A predição por si só apenas permite quantificar o grau de similaridade entre a predição e o dado real observado. O grau de similaridade, muitas vezes, é calculado pela correlação de *Pearson*, por exemplo, entre uma série temporal de um *voxel* ou um padrão espacial predito e observado entre *voxels* numa determinada região (ABDEL-GHAFFAR et al., 2015; KAY et al., 2008; NISHIMOTO et al., 2011). Embora a predição por um modelo de codificação já permita a comparação do desempenho do modelo entre regiões, do ponto de vista prático, uma aplicação de decodificação pode trazer informações adicionais e permite o desenvolvimento de tecnologias de “*leitura cerebral*” (NASELARIS et al., 2011). Motivado por essa perspectiva e pelo fato de que os estudos no domínio da música têm explorado pouco esse tipo de modelagem no domínio da música (ALLURI et al., 2013), este trabalho investiga a combinação de um modelo de codificação de características musicais com um modelo de decodificação com o objetivo de identificar músicas a partir da atividade cerebral. Utilizando uma analogia com um programa de TV, chamado “Maestro, qual é a música?”, no qual participantes devem identificar uma determinada música o mais

rápido possível, esse trabalho pergunta quanto tempo é necessário para identificar uma música com uma certa acurácia utilizando a atividade cerebral. Além do tempo, outras dimensões relevantes para o desempenho do modelo de decodificação são investigadas. Em específico, as seguintes hipóteses e perguntas são elaboradas:

- a) A acurácia aumenta com a duração do estímulo?
- b) A acurácia aumenta com a extensão do volume cerebral utilizado para a identificação?
- c) Quais são as regiões mais importantes para a identificação de músicas?
- d) Qual é a arquitetura anatômica-funcional das características acústicas?
- e) Quantas características acústicas o modelo necessita para um desempenho?
- f) O modelo é robusto para identificações de várias músicas?
- g) Existem diferenças na acurácia entre tipos ou estilos de músicas?

A investigação sistemática das dimensões mencionadas acima permite identificar parâmetros internos ou externos do modelo em relação ao seu desempenho. Todos os pontos mencionados são ingredientes importantes para um algoritmo e sistema em busca da otimização e do aperfeiçoamento de um modelo de decodificação. Com essa investigação e esse estudo detalhado, espera-se contribuir para desenvolvimentos futuros de tecnologias de leitura cerebral no domínio de música.

## **1.4 Organização do texto**

O texto é organizado em capítulos de fundamentação teórica, métodos, resultados e discussão. No capítulo 2, a fundamentação teórica começa com uma revisão da literatura sobre teorias de emoções e suas relações com música. Em seguida são citados estudos que investigaram respostas cerebrais à música. As últimas duas seções tratam da fundamentação metodológica. Primeiramente, os fundamentos da ressonância magnética e a utilização dessa tecnologia para a medição da atividade cerebral são tratados. E logo após, estudos que utilizaram essa técnica junto com modelos

matemáticos de codificação e decodificação para investigação do processamento visual, auditivo e da música.

O capítulo 3 apresenta o protocolo experimental elaborado nesse trabalho e a aquisição de dados. O protocolo é baseado em estudos revisados no capítulo anterior que adquiriram uma grande quantidade de dados em poucos participantes. Os vários passos de pré-processamento do dado são detalhados na seção 3.4. Nas seções seguintes a metodologia da análise principal é apresentada. No capítulo 4, os resultados das escalas comportamentais e respostas fisiológicas são apresentados nas seções 4.1 e 4.2. A seção 4.3 apresenta resultados sobre as primeiras abordagens de modelos de codificação aplicados nos primeiros cinco sujeitos adquiridos. Trata-se de análises comparativas com o objetivo de caracterizar o dado em relação a diversos parâmetros da modelagem. A análise principal, a combinação do modelo de codificação e decodificação para a identificação de músicas a partir da atividade cerebral é apresentada na seção 4.4. Parte dessa análise foi inspirada pelo já mencionado show de TV “Maestro, qual é a música. Nesse caso o algoritmo de identificação assume o papel dos participantes e visa identificar a música ao longo do tempo a partir da atividade cerebral. Essa análise principal inclui várias análises complementares que são detalhadas e discutidas na mesma seção. A discussão e a conclusão são apresentadas nos capítulos 5 e 6.

## 2 Fundamentação teórica

### 2.1 Teorias de emoções (musicais)

Embora que não haja consenso entre pesquisadores sobre a definição de emoção, a origem da palavra talvez possa indicar o papel dela: emoção no latim vem de *emotio* = movimento forte/abrupto e *emovere* = mover (para fora), comover, agitar. Nessa linha, o pesquisador Panksepp coloca a emoção como causa principal e essência da vida (LEWIS; HAVILAND-JONES; BARRETT, 2008, cap. 4). As emoções positivas são o estímulo para continuar vivendo, e já as negativas podem levar à escolha da morte, caso prevaleçam na vida do indivíduo. Ele ainda argumenta que a vida não faz sentido sem alegria e tristeza, e que são exatamente as emoções que levam as pessoas à interação social e criar laços afetivos e afiliativos.

Emoções são mais facilmente e naturalmente expressas na postura corporal e na face. Já a expressão de emoções em palavras apresenta uma barreira, que provavelmente contribui para o desacordo da definição de emoção. Mesmo sabendo dessa principal dificuldade, K. R. Scherer (2005) elabora uma definição programática na tentativa de criar um consenso e reduzir discussões vagas com termos imprecisos. Ele apresenta a seguinte definição:

*“[Emoção ...] é uma sequência de mudanças relacionadas e sincronizadas em todos ou quase todos dos cinco subsistemas do organismo em resposta à avaliação de um estímulo externo ou interno que é de importância maior para o organismo.”*  
[traduzido de SCHERER (1987)]

Dessa forma, Scherer foca na realização de emoções em subsistemas como um processo que envolve diferentes funções e componentes. Todos esses componentes deveriam ser medidos em conjunto no mundo ideal de pesquisa. Alguns componentes podem ser medidos de forma objetiva, por exemplo, mudanças no sistema nervoso central, respostas fisiológicas e expressões faciais, enquanto outros precisam de um relato verbal ou uma avaliação informada pelo sujeito.

Tabela 2-1 – Modelo da emoção em componentes de processo seguindo SCHERER (2005).

<b>Função</b>	<b>Subsistema ou substrato</b>	<b>Componente</b>
Avaliação de eventos e objetos	Processamento de informação (SNC)	Cognição (avaliação)
Regulação do sistema	Manutenção (SNC, SNE, SNA)	Neurofisiologia (sintomas corporais)
Preparação e direção de ação	Execução (SNC)	Motivação (intenções de ações)
Comunicação de reação e intenção comportamental	Ação (SNS)	Expressão motora (expressão facial e vocal)
Monitoramento do estado interno e interação do organismo com ambiente	Monitoramento (SNC)	Sentimento subjetivo (experiência da emoção)

Além de descrever emoções pelos subsistemas e componentes envolvidos, também é possível distinguir diferentes tipos de emoções. De um lado, encontram-se emoções *utilitárias* e, do outro, emoções estéticas. As primeiras são emoções como raiva, medo, alegria, nojo, tristeza, vergonha e culpa. Elas têm impacto direto e consequências importantes para o estado do organismo. Dessa forma, podem influenciar ou pelo menos modular ações, por exemplo, evitar objetos ou outros animais, motivar a se aproximar ou recuperar algum laço social como, por exemplo, se desculpar por alguma situação. Já emoções *estéticas*, geralmente, não afetam ações diretamente e estão situadas no âmbito da apreciação de obras de arte, apresentações artísticas ou musicais ou da beleza intrínseca da natureza. Esse grupo de emoções, geralmente, é verbalizado como *admiração, sublimidade, grandeza, deslumbramento ou solenidade* (SCHERER, 2005). Embora emoções estéticas possam ser consideradas mais abstratas, elas se manifestam de forma concreta em respostas fisiológicas como, por exemplo, na alteração do batimento cardíaco, na respiração e ainda numa intensidade mais alta como arrepios (GABRIELSSON; WIK, 2003).

A divisão entre emoções utilitárias e estéticas é também uma questão importante para as emoções induzidas por música. Embora o fator estético seja mais presente na apreciação de música (SCHERER, 2004), outros aspectos da música podem ter uma

explicação com origem na evolução e, por isso, uma relação mais intrínseca com as emoções utilitárias (JUSLIN 2013). Na busca de revelar a lógica intrínseca das emoções evocadas pela música, JUSLIN (2013) tenta unir os dois tipos de emoções num mesmo modelo, chamado BRECVEMA, usando oito mecanismos: reflexo do tronco cerebral, encadeamento rítmico, condicionamento avaliativo, contágio, imagética visual, memória episódica, expectativa musical e juízo estético. Esses oito mecanismos representam diferentes níveis, começando pelo reflexo do tronco cerebral, que é acionado por características acústicas como um som alto não esperado. Enquanto esses reflexos não são treinados, eles despertam e, num segundo momento, dentro de um contexto musical, podem condicionar uma avaliação cognitiva que resulta em surpresa. Encadeamento rítmico é outra característica da música que pode diretamente manipular o batimento cardíaco, respiração, além de causar alerta e sincronizar movimentos do corpo entre indivíduos. Já o condicionamento avaliativo é um mecanismo de um nível mais alto, envolvendo as associações e lembranças do contexto original de uma música conhecida (condicionamento). Uma música pode evocar alegria somente porque foi escutada várias vezes em contextos de alegria. O mecanismo de contágio reflete o aspecto que a emoção expressa na música facilmente ressoa no indivíduo. Isso acontece, principalmente, com música vocal e também com instrumentos semelhantes à voz, como instrumentos de corda. Outro aspecto é a evocação de imagens pela música. Muitas vezes, pessoas associam imagens ou cenas relacionadas à música ou mapeiam a estrutura musical em construtos não-verbais e abstratos. É possível também evocar memórias episódicas pela música e trazer emoções associadas como saudade ou nostalgia. Especula-se que essas memórias podem ser tão intensas porque, em parte, as respostas fisiológicas devem ser memorizadas (LANG, 1979).

### ***2.1.1 Modelos dimensionais e categorias de emoções***

Independente da música, emoções foram descritas cientificamente por Darwin pela expressão facial e corporal (DARWIN, 1872). Ele classificou emoções em categorias discretas, chamadas de emoções básicas. Cada emoção básica, como p.ex. medo, raiva, tristeza, alegria e nojo, é um produto da evolução em resposta a eventos e desafios do indivíduo. PLUTCHIK (1970) continuou a ideia de emoções básicas e criou quatro pares de emoções básicas opostas: raiva – medo, alegria – tristeza, confiança – nojo e antecipação – surpresa. Outras emoções (otimismo, amor, admiração, culpa, etc.) podem ser explicadas como mistura de emoções básicas.

Originalmente, as emoções básicas foram propostas como componentes ou dimensões independentes. RUSSELL (1980) tentou estruturar a relação entre essas emoções e usou métodos quantitativos para analisar uma série de experimentos nos quais os participantes tiveram de posicionar 28 adjetivos afetivos ao redor de um círculo. Os resultados revelaram as duas principais dimensões valência (agradável-desagradável) e grau de alerta que explicaram a maior variância dos posicionamentos. Este modelo dimensional está sendo utilizado em muitos estudos como referência, já que o modelo permite que qualquer emoção possa ser interpretada como uma combinação linear dessas duas dimensões. Num estudo mais recente, o mesmo grupo de pesquisadores propôs que cada dimensão tenha seu correlato neurofisiológico (POSNER, 2008).

Embora as duas dimensões de valência e alerta expliquem a maior variância entre categorias de emoções, elas não capturam sutilezas de emoções musicais. Isto foi comprovado pelo estudo de ZENTNER et. al (2008), que analisou termos emocionais associados a vários tipos de músicas. Respostas de mais de 1000 participantes foram avaliadas em vários experimentos. A análise de fatores revelou nove dimensões que possuem uma especificidade maior do que modelos de emoções básicos. As nove dimensões foram agregadas numa escala chamada GEMS (*Geneva Emotional Musical Scale*) representando encantamento, transcendência, ternura, nostalgia, serenidade, força, alegria, tensão e tristeza (ZENTNER; GRANDJEAN; SCHERER, 2008).

### **2.1.2 Emoções e alterações fisiológicas**

Existem várias teorias de emoções que debatem a importância e hierarquia dos componentes supostamente necessários para senti-las (KANDEL; SCHWARTZ; JESSELL, 2000, cap. 50). A teoria de James-Lange caracteriza emoções como respostas cognitivas a mudanças fisiológicas (JAMES, 1884). Numa teoria posterior, Schachter e Singer destacaram a importância do estado cognitivo, mostrando que a expectativa cognitiva pode modificar a emoção sentida (SCHACHTER; SINGER, 1962). Damasio estendeu essa teoria mais ainda, sugerindo que emoções são construtos ou histórias que o cérebro constrói para explicar as sensações fisiológicas. A teoria de Cannon-Bard revidou a teoria de James-Lange, mostrando que o córtex e estruturas subcorticais, em parte, são suficientes para criar reações emocionais como raiva, mesmo quando o sistema nervoso simpático aferente é removido. Dessa forma, Cannon e Bard inverteram



a hierarquia sugerindo que a origem das emoções esteja em estruturas subcorticais que, em seguida, alteram o estado fisiológico. Teorias mais recentes elaboram mais sobre os componentes corticais e subcorticais e as conexões (LEWIS; HAVILAND-JONES; BARRETT, 2008, cap. 4), mas todas elas têm em comum que emoções são, de alguma forma, acompanhadas por alterações fisiológicas.

As alterações fisiológicas se manifestam, entre outros, na respiração, batimento cardíaco, tensão muscular, sudorese e sistema vasomotor. A medição desses fenômenos periféricos é importante por dois motivos: primeiro, ela pode demonstrar efeitos diretos da experiência emocional durante um experimento; segundo, alterações fisiológicas podem afetar também a medida do sinal BOLD, conhecido como ruído fisiológico. Principalmente, os ciclos cardíaco e respiratório têm efeito direto na magnetização, e, por isso, podem ser confundidos com o sinal BOLD (KASPER, 2013). Por exemplo, a sístole e a diástole alteram a pressão nas artérias, o que, subsequentemente, altera o volume delas. Essa alteração de volume, em seguida, é compensada por um fluxo do fluido cerebrospinal (CSF). Da mesma forma, durante cada ciclo respiratório, o volume dos pulmões desloca levemente o tecido do corpo e, subsequentemente, altera o campo magnético. Enquanto esse efeito tem impacto mais global na codificação do espaço da imagem, o ciclo respiratório também afeta diretamente a oxigenação do sangue. Essa influência é relacionada à fase do ciclo respiratório e, por isso, não deve ser confundida com o sinal BOLD, que reflete o consumo de energia metabólica.

## **2.2 Respostas cerebrais a música**

Vários estudos investigaram a resposta cerebral à música de diversas formas. Nessa seção, apenas alguns estudos mais relacionados com o presente projeto serão recapitulados. Os correlatos cerebrais de emoções evocadas por música foram revisados recentemente por KOELSCH (2014). Várias estruturas são recrutadas pela música, e a integração e comunicação entre essas estruturas formam uma percepção emocional. O caminho do processamento neural inicia na cóclea. Ela envia informações pelo nervo auditivo para os núcleos cocleares, situados no tronco cerebral, que, em seguida, comunicam com o colículo inferior do mesencéfalo e outros núcleos no tronco cerebral. Já esse nível causa alerta e estimula primeiras reações como sobressalto e outros

movimentos. O corpo geniculado medial, um núcleo situado no tálamo, recebe projeções do colículo inferior e possui projeções bidirecionais com o córtex auditivo. A amígdala tem um papel central nesse meio do caminho. Ela possui diferentes subdivisões e estabelece conexões diretas com várias estruturas, como o córtex auditivo, o corpo geniculado medial, o núcleo acumbente, o córtex orbito-frontal e a ínsula. Por isso, ela é considerada um nó central, que pode modular, iniciar, manter e terminar respostas emocionais. Em vários estudos, uma relação com valência positiva e negativa foi observada. Além disso, ela tem um papel importante para consolidação de memória. Outra estrutura, associada a estímulos prazerosos e agradáveis, é o núcleo acumbente, situado no estriado ventral. Vários estudos mostraram que música é capaz de ativar essa região, principalmente em momentos de respostas emocionais intensas (BLOOD; ZATORRE, 2001). Especialmente, a conexão funcional do núcleo acumbente com o córtex auditivo, córtex orbito-frontal e pré-frontal ventro-medial mostrou-se diretamente relacionada ao prazer evocado pela música (SALIMPOOR et al., 2013). Outra estrutura envolvida com o processamento de música é o hipocampo. Principalmente, músicas conhecidas que trazem lembranças podem recrutá-lo. Mas também já foi observada a ativação por emoções de alegria, ternura, serenidade e tristeza (KOELSCH, 2014). Por último, regiões motoras e o córtex cingulado podem ser recrutados pela música. As regiões motoras podem, por exemplo, representar vontade de dançar ou movimentos imaginados que acompanham a música. As mudanças fisiológicas, evocadas pela música, ativam o córtex cingulado, que é responsável pelo monitoramento e consciência interoceptiva.

Está consolidado que música é capaz de evocar emoções e recrutar várias estruturas cerebrais. Porém, qual é o mecanismo exato e o que é necessário para gerar as emoções? Posto de outra forma, quais características musicais são responsáveis por evocar emoções? Um estudo de CHAPIN et al. (2010) tenta responder exatamente a essa pergunta. Num experimento bem controlado, eles apresentam a mesma música em duas versões: uma versão tocada pelo artista com expressão, incluindo flutuações temporais e variações de volume, e outra mecânica produzida pelo computador, que tocava cada nota com o tempo exato e com a mesma intensidade. Apenas a versão interpretada pelo artista conseguiu evocar as estruturas cerebrais responsáveis pelas emoções. As variações no tempo e na intensidade estavam altamente correlacionadas com alerta subjetivo dos participantes. Os autores concluíram que essas variações são

essenciais para produzir as emoções. Eles explicam que existe uma interação dinâmica entre o estímulo musical e o processamento neural. Essa interação gera expectativas da próxima nota. Quando essa expectativa interna neural é violada, a discrepância será registrada em estruturas profundas, causando alerta e reações emocionais. A violação pode até funcionar como recompensa, já que um evento não esperado contém muita informação que pode ser utilizada para melhorar o modelo de expectativa (KOELSCH 2014).

Voltando a um nível mais básico e inicial, pode-se perguntar como o processamento neural de estímulos musicais se compara com o processamento de outros estímulos auditivos. Essa questão pode parecer simples, porém, na prática, não é fácil responder. PRICE et. al (2005) resumem vários estudos que compararam estímulos linguísticos com estímulos musicais. Foi constatado que linguagem e música compartilham regiões cerebrais em diferentes etapas do processamento auditivo. Em detalhe, a região ativada pelo contraste de fala versus ruído é a mesma do contraste de sons versus ruído. Isto ainda é um resultado esperado, já que, nos dois casos, compara-se um estímulo auditivo com silêncio. Porém, em contrastes mais específicos, os autores também observaram grande sobreposição. O contraste entre FALA e NÃO-FALA mostrou regiões ativadas no giro temporal superior anterior e posterior ao giro do Heschl. Essas mesmas regiões foram ativadas para a comparação MELODIA versus TONS. Os autores concluíram que, no nível macro-anatômico, as áreas do processamento neural de estímulos de linguagem e de música são, em grande parte, compartilhadas. Porém, as áreas compartilhadas podem possuir algumas diferenças e especializações. ZATORRE et. al (2002) mostram uma preferência do giro temporal esquerdo para características temporais que são importantes para o entendimento fonético. Já o giro temporal direito possui maior capacidade em detectar diferenças sutis em padrões espectrais.

O processamento de características musicais de baixo nível também mostra de forma similar estruturas compartilhadas e sub-regiões mais específicas. Mudanças de altura de tons, realizadas por diferentes atenuações de múltiplas da frequência fundamental, ativaram o giro de Heschl, assim como mudanças de tons em intervalos de semitons. Porém, existe maior ativação para a altura posterior ao giro de Heschl, enquanto a mudança em semitons recrutou uma área mais anterior (WARREN et al., 2003).

## 2.3 Ressonância magnética

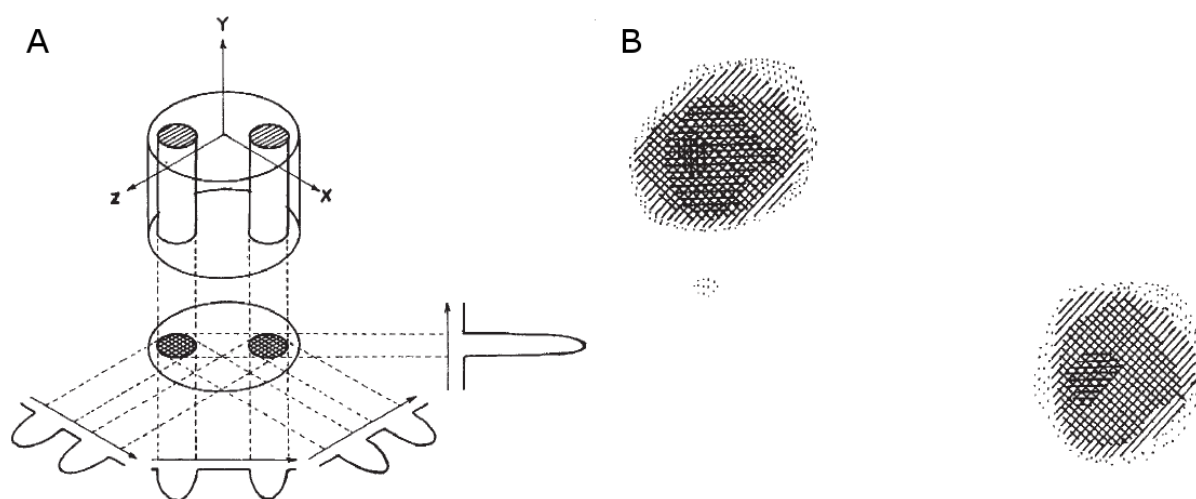
### 2.3.1 História de RM

A ressonância magnética (RM) aproveita das características magnéticas inerentes de partículas nucleares que foram descritas e descobertas nos anos 1920-1935 durante a elaboração da mecânica quântica. Em 1924-1925, os primeiros trabalhos sugerem a existência de dois possíveis estados de elétrons, chamado *spins*, que, por sua vez, possuem um momento angular e magnético (PAULI, 1924; UHLENBECK; GOUDSMIT, 1925). Interessantemente, um experimento de Stern e Gerlach demonstrou um momento magnético de átomos de prata antes da teoria do momento magnético dos elétrons (GERLACH; STERN, 1922). Nesse experimento, átomos de prata foram projetados ao longo de um magneto até chegar num detector, que registrou o desvio da trajetória dos átomos causado pelo magneto. O detector mostrou apenas um ponto superior e outro inferior ao centro da trajetória indicando que existe um momento magnético com dois possíveis estados. O experimento de Gerlach e Stein serviu como base da demonstração da ressonância magnética por RABI (1939): ele mostrou que átomos podem absorver energia de um campo magnético, no caso em que a frequência estiver em ressonância com o momento magnético e angular do átomo. Essa frequência de ressonância depende tanto da característica do átomo como do campo magnético externo. Formalmente, a frequência com máxima ressonância  $\omega$ , a frequência de Larmor, é proporcional à constante giromagnética  $\gamma$  do átomo e ao campo magnético externo  $B$ :  $\omega = \gamma B$ .

Mais tarde, em 1946, Purcell e Bloch confirmaram com diferentes abordagens o efeito de ressonância magnética em substâncias sólidas. Já nessa época, o experimento de Bloch utilizou os mesmos componentes de sistemas atuais de ressonância magnética: um campo magnético estático, uma bobina de transmissão de energia de radiofrequência e uma bobina de detecção. Nota-se que, antes de a ressonância magnética nuclear ser aplicada para aquisição de imagens, a tecnologia era utilizada para analisar estruturas químicas (HUETTEL; SONG; MCCARTHY, 2004).

Até o final da década de 1960, a RM media apenas um tecido como um todo, ou seja, não tinha informação espacial. Nessa época, DAMADIAN (1971) descobriu que é possível diferenciar tecidos biológicos, como, por exemplo, tumores malignos de

tecidos normais, com a RM. Isso motivou os próximos passos para que se obtivesse informação espacial de uma amostra. A primeira imagem 2D de duas amostras de água foi obtida com sucesso por LAUTERBUR (1973) por um processo iterativo de um gradiente do campo magnético de uma dimensão (confere Figura 2-1). Alguns anos depois, esse processo iterativo foi substituído pela aplicação da transformada de Fourier em 2D e 3D (KUMAR; WELTI; ERNST, 1975). A primeira imagem do corpo humano foi adquirida por DAMADIAN et. al (1977). A aquisição levou 4 horas para obter uma imagem do tórax com 106 pixels. Em paralelo, a técnica foi aprimorada pela imagem eco-planar (EPI), que permite a leitura de uma imagem 2D após uma única excitação. Essa técnica é fundamental até hoje para realizar exames num tempo factível e também abriu o caminho para a investigação das funções do cérebro.



*Figura 2-1: Primeira imagem obtida com RM. A: Ilustração da amostra e do processo da reconstrução da imagem. Os gradientes foram aplicados em quatro orientações diferentes (indicadas pelas setas). B: Reconstrução da imagem obtida pela projeção dos quatro gradientes. [Imagem originais da publicação de (LAUTERBUR, 1973)]*

### **2.3.2 Sistema de RM**

Uma máquina de RM possui vários componentes: o magneto do campo principal, um sistema de gradientes  $x$ ,  $y$  e  $z$  para codificação do espaço das imagens em 3D, bobinas de transmissão e bobinas de recepção. Geralmente, as máquinas utilizam um eletromagneto cilíndrico com super-condutância. Nas camadas interiores do cilindro, são localizados as bobinas de gradientes e de transmissão. Além da bobina de recepção de corpo integrado ao cilindro, existem bobinas específicas e portáteis para a

região examinada (abdômen, pescoço, crânio etc.). Duas salas separadas da sala do magneto fazem parte do sistema como um todo: a sala técnica, com computador (reconstrutor das imagens), microcontroladores, amplificadores e sistema de refrigeração, e a sala de comando, com um computador para realização dos exames.

### 2.3.3 Formação de imagem

O conceito central para a formação de imagem por RM é a aplicação de gradientes. Um gradiente é um campo magnético que varia ao longo de uma certa dimensão ( $x$ ,  $y$  ou  $z$ ). Essa variação controlada permite estabelecer uma correspondência da frequência de Larmor com cada posição ao longo de uma dimensão na imagem. Para obter uma imagem em duas dimensões, acrescenta-se um segundo gradiente de fase. Com isso, a codificação de uma imagem 2D é dada pela correspondência de cada posição na imagem  $x$ ,  $y$  com certa frequência e fase. Essa correspondência é matematicamente descrita pela transformada de Fourier 2D:

$$S(t) = \int_x \int_y M(x, y) e^{-i2\pi k_x(t)x} e^{-i2\pi k_y(t)y} dx dy \quad (1)$$

onde  $k_x(t)$  e  $k_y(t)$  são dados pela trajetória no *espaço-k* de frequência e fase, ou seja, pela sequência dos gradientes  $x$  e  $y$ ;  $M(x, y)$  é o sinal de magnetização na posição  $x$  e  $y$ ;  $S(t)$  é o sinal bruto obtido pela bobina de recepção. O sinal bruto obtido no *espaço-k* é reconstruído com a inversa da transformada de Fourier para obter as imagens 2D finais. Acrescenta-se um gradiente de seleção de corte no eixo  $z$  combinado com a radiofrequência correspondente para obter um volume completo em 3D.

### 2.3.4 Sinal BOLD

A magnetização da soma dos *spins* pode ser capturada pela bobina de recepção. Essa magnetização é zero antes da excitação dos prótons porque a precessão dos *spins* é alinhada com o campo magnético. O pulso de excitação pela bobina de transmissão muda o ângulo da precessão dos *spins*, que resulta numa magnetização transversal. Ao longo do tempo, essa magnetização transversal diminui devido ao processo chamado relaxação. A relaxação é influenciada pelo componente longitudinal e transversal. O componente longitudinal, chamado de  $T_1$ , pode ser explicado pela recuperação do

estado inicial devido à perda de energia. O componente transversal é resultado da perda de fase entre os *spins*. Essa perda de coerência de fase é causada por interações entre *spins* ( $T_2$ ) e por inhomogeneidades do campo magnético. A inhomogeneidade do campo muda a frequência da precessão, o que causa uma perda adicional de coerência de fase ( $T_2^*$ ). Esse efeito é aproveitado na RM funcional: sangue com hemoglobina oxigenada é diamagnético, resultando numa homogeneidade maior, conseqüentemente num decaimento do sinal  $T_2^*$  menor, e, por isso, num sinal maior (HUETTEL; SONG; MCCARTHY, 2004).

## 2.4 Codificação e decodificação respostas cerebrais

Nesta seção, serão recapitulados estudos com abordagens de codificação e decodificação. Trata-se de uma parte mais técnica voltada a técnicas de análise. Uma visão geral dessas análises é dada na

*Tabela 2-2.*

### 2.4.1 Domínio visual

Um dos primeiros estudos de codificação e decodificação de estímulos ocorreu no domínio visual (KAY et al., 2008). Os autores apresentaram, em várias sessões de RMf, 1.750 imagens de fotografia com intervalos de 3 s entre elas. Depois de ter utilizado essas sessões para treinar o modelo, eles registraram a resposta cerebral para um novo conjunto de 120 imagens. Os 500 melhores *voxels* do treino nas áreas visuais V1, V2 e V3 foram utilizados para, então, identificar qual das 120 novas imagens o sujeito estava vendo no momento. A análise foi feita em 2 sujeitos individualmente. Enquanto, nas sessões de treinamento, cada imagem foi apresentada duas vezes, as novas imagens foram apresentadas 13 vezes. A estimativa da resposta BOLD envolve vários passos. Primeiro, o formato da resposta hemodinâmica é estimado por *voxel*. Esse formato é uma combinação de funções básicas de seno e cosseno com 3 a 4 ciclos se estendendo de 1 s a 16 s depois do início da apresentação do estímulo. Enquanto o formato é assumido igualmente para todas as imagens, a amplitude da resposta é estimada individualmente por imagem. Essas estimativas transformam o dado bruto num valor representativo da atividade por *voxel* e por estímulo. Baseado nesses valores,

o mapeamento entre estímulo e resposta cerebral é realizado. Para isto, cada imagem foi descrita num modelo hierárquico de *wavelets* de Gabor. Essa descrição envolveu, no total, 2.730 parâmetros, que representam localização, orientação e frequência espacial do campo receptivo. A estimativa do campo receptivo por *voxel* foi realizada com um método iterativo (método do gradiente). Esse método minimiza o erro de forma iterativa utilizando como critério de término o erro sobre um dado não utilizado para estimar a direção do gradiente. O algoritmo termina assim que erro sobre o dado de teste não diminui mais. No final, a solução dessa estimativa representa o campo receptivo do *voxel*, ou seja, as preferências de frequências espaciais e de orientação das *wavelets* de Gabor.

Com essa abordagem, as respostas dos melhores 500 *voxels*, selecionados durante o treino, formaram a base para a identificação de imagens. Para os dois sujeitos, 110 e 86 de 120 imagens foram identificadas corretamente, resultando numa acurácia de 92% e 72% para sujeito 1 e 2, respectivamente. Essas acurácias foram obtidas com a resposta média sobre 13 apresentações da mesma imagem para aumentar a razão entre sinal e ruído. Quando aplicado na resposta única por apresentação de imagem, as acurácias caíram para 51% (834/1620) e 32%.

Na área de modelagem do sistema visual, ainda houve outros estudos do mesmo grupo: NISHIMOTO et al. (2011) investigou a codificação neural de estímulos visuais dinâmicos. O modelo de codificação é uma versão estendida das *wavelets* de Gabor, englobando o tempo como dimensão adicional. Para estimar os parâmetros do modelo, foram coletados dados de RMf durante a apresentação de vídeos com a duração total de 7.200 s. Um outro conjunto contendo 9 vídeos de 1 min foi apresentado separadamente com tempo total de 5.400 s (cada vídeo foi apresentado 10 vezes). Três diferentes modelos foram testados: um modelo estático, um modelo englobando a dimensão temporal, detectando energia de movimento não direcionado, e um terceiro modelo detectando direcionalidade do movimento. De 4.500 *voxels* iniciais nas áreas visuais V1, V2, V3, V3A e V3B, os melhores 2.000 *voxels* foram selecionados, baseado numa validação cruzada interna no dado de treino que dividiu o dado em partes de 90% para estimar o modelo e 10% para validar a predição em cada *voxel* separadamente. A divisão do dado foi feita em blocos de 50 s para garantir a independência entre o dado de treino e de validação. A média da correlação de predição sobre os três sujeitos e os melhores 2.000 *voxels* foi 0,24, 0,39 e 0,4 para os três modelos respectivamente,



mostrando que a modelagem com a dimensão temporal é superior à modelagem estática. Baseado nessa estimativa do modelo, os autores conseguiram identificar o conteúdo dos 9 vídeos utilizando o conjunto dos 2.000 *voxels* ao mesmo tempo. O sinal multivariado predito pelo conteúdo atual coincidiu com o sinal medido do mesmo instante dentro de uma margem de  $\pm 1$  s em 95% do tempo para um dos três sujeitos. Os autores mostraram ainda que a modelagem é tão específica que, mesmo dentro de um conjunto de 1.000 vídeos, o trecho correto de vídeo é identificado em 75% das vezes para cada um dos três sujeitos. Por fim, os autores tentaram reconstruir o estímulo visual utilizando uma base com 18 milhões de trechos de vídeos de 1 s. Os 100 trechos mais prováveis de ter evocado o sinal observado nos 2.000 *voxels* entraram na reconstrução. Uma simples média desses trechos obteve a melhor acurácia de reconstrução. A acurácia foi calculada como correlação entre estímulo original e estímulo reconstruído no espaço das *wavelets* de Gabor com a dimensão temporal englobada. Essa acurácia estava em torno de 0,3 para os três sujeitos e resultou numa reconstrução visualmente semelhante ao estímulo original.

Além dos dois estudos resumidos acima, pesquisadores do mesmo grupo investigaram com abordagens semelhantes várias outras questões: a reconstrução de estímulos visuais a partir da combinação de informações visuais, semânticas e contextuais (NASELARIS et al., 2009). Enquanto o modelo restrito para informações visuais mostrou-se superior nas áreas V1 e V2, o modelo com informação semântica, representado por 23 categorias hierárquicas, atingiu uma predição melhor nas áreas V3. A questão da representação semântica é investigada especificamente num estudo posterior. HUTH et al. (2012) tentam decifrar o código neural de representações semânticas de 1.705 categorias de objetos. Os autores mostram que objetos são representados num espaço contínuo, que reflete as distâncias e proximidades entre objetos e categorias. A representação de similaridades entre objetos provavelmente leva a um armazenamento mais eficiente e compacto, já que objetos similares compartilham algumas características em comum. Em outro estudo mais recente, ÇUKUR et al. (2013) investigam como o espaço semântico é influenciado pela atenção. Os participantes do estudo assistiram a vários vídeos e tiveram que focar a atenção ora em humanos ora em veículos. Observou-se uma mudança no mapeamento do espaço semântico no sentido que objetos associados à categoria alvo obtiveram um aumento de

peso neste mapeamento. Os autores interpretaram esse resultado como influência dinâmica pela atenção na representação neural.

Os estudos da área visual têm em comum uma resolução relativamente alta em torno de 2-3 mm por dimensão. O número de sessões e duração total do experimento, muitas vezes adquirido com TR de 1 s ou 2 s, também se destaca. Essa grande quantidade é necessária para se ter um equilíbrio na estimativa de modelo, já que o número de variáveis é muito grande (geralmente maior que 1.000 parâmetros). Por isso, também é necessário aplicar alguma forma de regularização em conjunto com a validação cruzada para obter estimativas robustas e evitar o sobreajuste.

#### **2.4.2 Domínio auditivo**

No domínio auditivo, também existem vários estudos que investigam representações de sons. Muitos estudos focaram na tonotopia, ou seja, na representação de frequências no córtex auditivo (HERDENER et al., 2013; SAENZ; LANGERS, 2014; SCHÖNWIESNER et al., 2014; STRIEM-AMIT; HERTZ; AMEDI, 2011; THOMAS et al., 2015). Embora todos os estudos tenham identificado gradientes de frequência ao longo do giro de Heschl, a orientação exata dos gradientes ainda vem sendo discutida. Da mesma forma, ainda não há consenso sobre a separação clara entre áreas primárias e secundárias (BAUMANN; PETKOV; GRIFFITHS, 2013; HUMPHRIES; LIEBENTHAL; BINDER, 2010; LANGERS; VAN DIJK, 2012; MOEREL, 2013). Parecido com o sistema visual, existem representações de uma ordem mais alta no córtex auditivo. Tanto localização de fontes de sons como categorização de sons foram evidenciadas por vários estudos. Por exemplo, LEAVER e RAUSCHECKER (2010) identificaram representações distintas para vozes humanas e instrumentos musicais. Mesmo assim, a relação e representação de sons complexos, além da representação tonotópica, estão longe de serem compreendidas. Não se sabe, ainda, quais são os caminhos de processamento neural num estágio depois da tonotopia, quais características são processadas em paralelo ou de forma sequencial e como, finalmente, uma percepção acústica completa de um som é realizada (MOEREL, 2013).

Uma tentativa em direção ao entendimento do processamento neural além da tonotopia é um estudo de SANTORO et al. (2014). Os autores construíram alguns modelos que descrevem sons num espaço frequência-temporal. Enquanto o modelo de tonotopia interpreta respostas neurais apenas em relação a frequências constantes, os

modelos utilizados no estudo incluíram uma descrição mais completa. Mais em detalhe, além de responder a uma frequência em si (em humanos na faixa de 20 Hz a 20 kHz), as respostas neurais poderiam ser moduladas de forma temporal ou espectral. A modulação temporal pode ser vista como a modulação em amplitude numa certa frequência moduladora (no estudo as frequências escolhidas para modulação temporal eram de 1, 3, 9 e 27 Hz). Por exemplo, alguns neurônios poderiam responder melhor a um tom de 400 Hz, se a amplitude dele alterna três vezes por segundo, ou seja, se o volume aumenta e diminui em ciclos de 3 Hz. Por outro lado, a modulação espectral permite que neurônios respondam melhor a múltiplas frequências, por exemplo, oitavas ou harmônicas de uma frequência fundamental (no estudo com os valores de 0,5, 1, 2 e 4 ciclos por oitava). Baseados nessas duas principais modulações, os autores construíram vários modelos: modelos com apenas uma das duas modulações e um modelo com a combinação das duas modulações. Cada um desses modelos ainda foi subdividido em modelos específicos para certas faixas de frequências ou inespecíficos, isto é, o valor médio das respostas sobre todas as frequências perceptíveis ao ouvido.

Os modelos foram aplicados a dados de RMf adquiridos durante dois experimentos em máquinas de 3T e 7T. Os participantes escutaram 60 e 168 sons no total com duração de 1 s, oriundo de vozes humanas, animais, instrumentos musicais, da natureza e de ferramentas. No experimento na 7T, os sons foram apresentados três vezes, e dados de treino (144 sons) e teste (24 sons) foram adquiridos em sessões separadas. No experimento na 3T, os 60 sons foram apresentados nove vezes (3 vezes por *run*). A estimativa do modelo seguiu de forma similar a abordagem de KAY et al. (2008). Primeiro, o formato da resposta hemodinâmica foi determinado por *voxel* pela resposta média sobre todos os estímulos. Em seguida, os pesos das parametrizações foram obtidos por uma regressão de *ridge*. Em vez de identificar o valor do parâmetro da regularização mais adequado por validação cruzada, os autores inspecionaram as mudanças dos pesos iniciais ao longo do aumento do valor da regularização. O melhor equilíbrio entre uma estimativa robusta e demanda computacional foi encontrado no término precoce, quando o aumento do valor da regularização resultou em mudanças menores do que 20% em todos os parâmetros. Além disso, uma proporção adequada entre número de parâmetros e número de pontos disponível foi obtida por utilizar apenas três faixas de frequências básicas no experimento na 3T. Isso resultou num modelo com 48 parâmetros (modelo combinado por 4 modulações temporais e 4

modulações espectrais e 3 frequências) para 60 pontos. Já para o experimento na 7T, eles representaram as frequências básicas por 8 faixas, obtendo 128 parâmetros para 144 sons. O desempenho dos modelos foi avaliado por um esquema de identificação dos sons de teste. A identificação dos 60 e 24 sons foi representada por um ranqueamento, já que o som correto, muitas vezes, atingiu somente uma predição alta e não a predição maior entre todos os sons testados. Dessa forma, o ranqueamento representa de forma mais global a capacidade dos modelos, indicando o desempenho superior ao nível de chance quando ultrapassa 50% no ranqueamento. Os autores destacaram no resultado principal que o modelo combinado com duas modulações era superior aos restantes modelos com apenas uma das duas modulações e ao modelo tonotópico. Além disso, a identificação no experimento da 7T com 144 sons era melhor do que no experimento de 3T com 60 sons para todos os modelos, o que destaca a importância de se ter um número maior de exemplos para estimar o modelo. O número maior de exemplos provavelmente permite capturar variações maiores nos parâmetros do modelo. Além disso, os autores mencionaram o fato de o contraste funcional ser maior na 7T e da resolução maior utilizada nesse experimento. Na discussão final, os autores discutem alguns pontos que podem explicar por que as acurácias de identificação são menores no domínio auditivo do que no domínio visual. Por um lado, a resposta neural a sons é diminuída pelo barulho da máquina. Para reduzir esse efeito, os autores incluíram um intervalo de silêncio entre volumes, o que, por outro lado, diminui o número de sons possíveis no protocolo. Apesar de os resultados não terem alcançado o mesmo patamar dos estudos na área visual, esse estudo foi o primeiro a mostrar que técnicas de codificação permitem a comparação de modelos computacionais de processamento auditivo, inspirados por modelos biológicos e aplicados a sons reais.

### **2.4.3 Domínio da música**

O mapeamento de características musicais na atividade cerebral com técnicas de codificação e decodificação ainda está no início das investigações. Um dos primeiros estudos investigou a resposta cerebral na sua evolução temporal a músicas completas (ALLURI et al., 2013). Onze participantes escutaram um *medley* dos Beatles (*Abbey Road*) e outros doze participantes um *medley* composto por diferentes gêneros (clássica, jazz, pop). Desses dois *medleys* de 15 min de duração cada, 25 características musicais foram extraídas com a MIRtoolbox (LARTILLOT; TOIVAINEN, 2007). A maior parte das características descreve alguma informação espectral e são extraídas ao longo de

tempo em janelas de 25 ms. Algumas outras características são de uma ordem mais alta, e descrevem, por exemplo, a clareza da marcação do ritmo e do modo. Para utilizar essas características na modelagem da resposta BOLD, elas foram primeiramente convoluídas com uma resposta hemodinâmica padrão e, em seguida, reamostradas a 0,5 Hz para coincidir com a amostragem do sinal de RMf. Como último passo antes da análise principal, a média do sinal BOLD foi calculada por grupo de participantes, e as características musicais de cada música foram transformadas para componentes principais para evitar problemas de colinearidade. Nove componentes principais por música entraram na regressão linear múltipla. Os pesos dos componentes foram estimados por música e por *voxel* e, em seguida, utilizados para prever a atividade cerebral da outra música (validação cruzada entre estímulos e grupo de participantes). Com essa abordagem, os autores identificaram os maiores *clusters* com previsões significantes no giro temporal superior. A música dos Beatles obteve mais *voxels* significantes do que o *medley*. Além disso, o hemisfério direito foi melhor predito do que o hemisfério esquerdo para a música dos Beatles, enquanto o contrário foi observado para a música de *medleys*. Os autores ainda investigaram a generalização do método numa terceira música, que foi adquirida num estudo anterior. A validação cruzada com essa música também mostrou uma leve preferência para o giro temporal superior direito. Outras regiões, como o precuneus, giro frontal superior, área motora suplementar, opérculo rolândico, giro pós-central e parietal, também foram preditas com acurácias significantes, embora com valor estatístico mais baixo. Em resumo, esse estudo mostrou que a atividade cerebral pode ser predita a partir de um conjunto de características musicais e que essa previsão pode ser generalizada entre grupos de participantes e gêneros musicais.

Em outro estudo, TOIVAINEN et al. (2014) inverteram a direção de codificação. Em vez de o conjunto de características musicais prever a atividade de um *voxel*, o padrão espacial entre *voxels* foi utilizado para explicar um componente musical. Quinze participantes escutaram o lado B do álbum *Abbey Road* dos Beatles durante 16 min. A extração das características musicais seguiu o mesmo procedimento de ALLURI et al. (2012, 2013). Na análise principal, entraram seis componentes principais após rotação pelo método *varimax*, previamente validados na percepção subjetiva como *completude*, *brilho*, *atividade*, *complexidade timbral*, *clareza de ritmo* e *clareza de modo* em ALLURI et al. (2012). Para testar quão bem cada um dos seis

componentes pode ser predito pela atividade cerebral, os autores treinaram um modelo com uma regressão *lasso*, e validaram o modelo entre sujeitos. Para cada iteração de treino e teste, um participante foi reservado para a validação. Dos demais participantes, os *voxels* mais robustos foram escolhidos pela correlação temporal entre pares de sujeitos. O número de *voxels* foi variado de 1/4, 1/8, 1/16, 1/32, ou 1/64 do número total dos *voxels*. A média sobre os participantes foi calculada *voxel a voxel*, e, com base nessa média, uma análise de componentes principais foi utilizada para representar o padrão espacial entre *voxels*. A regressão *lasso*, enfim, estimou os pesos para o conjunto dos componentes espaciais. A validação consistiu em aplicar os pesos estimados para prever a evolução temporal de cada componente musical. Além de variar o número de *voxels* na seleção dos *voxels* mais robustos, o número de componentes espaciais, ou seja, o número de variáveis da regressão variou de 1 a 100. Os melhores resultados foram obtidos para as proporções 1/8 a 1/64 dos números de *voxels* da pré-seleção e 2 a 32 componentes espaciais. A predição, medida pela correlação temporal, obteve valores moderados de 0,3 a 0,5 para cada componente musical, menos o componente *clareza de modo*, que ficou por volta de 0,2. Os pesos estimados na regressão foram transformados de volta para o espaço original dos *voxels* para visualizar as regiões mais importantes para a decodificação dos componentes musicais. Os mapas espaciais para *completude*, *brilho*, *atividade* e *complexidade timbral* mostraram *clusters* no giro de Heschl direito e, para *brilho*, *atividade*, *clusters* adicionais no giro temporal médio esquerdo. A *complexidade timbral* se destaca entre eles por ter *clusters* maiores por volta do giro de Heschl e serem simétricos bilaterais. A *clareza de ritmo* tem um padrão mais distribuído, com contribuições de cíngulo anterior e posterior e giro frontal inferior, indo na direção medial orbital. Por fim, os autores apresentam resultados de identificação de segmentos, que foram recortados com tamanhos iguais da música original. Eles variaram o número de segmentos a serem identificados de 2 a 10 e viram que a identificação em todos os casos foi bem além do nível de chance, passando limiares estatísticos obtidos por simulações. Em resumo, os autores conseguiram confirmar resultados de estudos anteriores com uma nova abordagem, que generalizou informações sobre diferentes participantes. Entre os pontos confirmados, se destacam a melhor predição para características timbrais de um nível mais baixo e as áreas mais importantes para a decodificação como os giros temporal superior, incluindo o giro do Heschl. A lateralização para o giro de Heschl direito e giro temporal medial também foi confirmada. No demais, algumas características obtiveram melhor acurácia com apenas

4% dos *voxels*, o que pode ser interpretado como uma representação econômica dessas características.

Tabela 2-2 – Relação de número de variáveis, pontos e modelos em publicações de codificação. *Parametrização: Wavelets de Gabor (WG), Wavelets de Gabor incluindo dimensão temporal (tWG), Etiquetas semânticas (ES), Filtros modulados por tempo e frequência (FMTF), Características musicais (CM) ; Formato da resposta hemodinâmica: Resposta por funções básicas (RFB), Resposta ao impulso finito (RIF), Função hemodinâmica de double gamma (2G); Estimador: Gradiente descendente com término precoce (GD), Regressão ridge (RR), Regressão lasso (RL), Regressão por componentes principais (RCP), Validação cruzada (VC)*

<b>1º Autor, Ano</b>	<b>Domínio</b>	<b>Número variáveis</b>	<b>Número pontos</b>	<b>Nº. de sujeitos</b>	<b>Modelo</b>	<b>Validação</b>
Kay, 2008	Visual	2.730	1.750 (imagens)	2	WG, RFB, GD	Identificação de 1 entre 120 imagens
Naselaris, 2009	Visual, Semântico	10.921 +23	1.750 (imagens)	3	WG+ES, RFB, GD	Reconstrução de 120 imagens comparando modelos com e sem ES
Nishimoto, 2011	Visual	6.555	7.128 (segundos de vídeo)	3	tWG, RIF de (3-6 s), RL	Identificação temporal e reconstrução de 9 min de vídeos
Huth, 2012	Semântico	1.705	3.600	5	ES, RIF (4, 6 e 8 s), RR	VC via bootstrap. Componentes na representação semântica
Çukur, 2013	Semântico, Atenção	935	900 (por condição)	5	ES, RIF (2, 4 e 6 s), RR	VC via bootstrap. Avaliação de mudança nos mapeamentos devido à atenção

Continuação da Tabela 2-3:

<b>1º Autor, Ano</b>	<b>Domínio</b>	<b>Número variáveis</b>	<b>Número pontos</b>	<b>Número sujeitos</b>	<b>Modelo</b>	<b>Validação</b>
Santoro, 2014	Auditivo	48 (3T) 128 (7T)	60 (3T) 144 (7T)	5	FMTF, RIF, RR	VC entre runs – Identificação do mesmo conjunto de estímulo (3T) Identificação em 24 sons distintos (7T)
Alluri, 2013	Música	9	400-450	11	CM, 2G, PCR	VC entre estímulos e grupo de participantes
Toiviainen, 2014	Auditivo	1-100	464	15	CM, 2G, RCP	VC entre sujeitos, Decodificação de CM



## 3 Métodos

### 3.1 Participantes

Participaram do experimento 6 sujeitos saudáveis (idade média: 30,8  $\pm$ 7,8 DP; 1 masculino) com experiência musical prática atual ou no passado (média de horas ativas: 4424  $\pm$ 4198 DP). Nenhum participante relatou algum problema auditivo. Todos os participantes escutaram as músicas antes de participar do estudo e informaram que eram capazes de evocar emoções com as músicas escolhidas pelos experimentadores, descritas a seguir. Um sétimo participante foi excluído depois da primeira sessão porque relatou dificuldades de evocar as emoções pretendidas.

### 3.2 Estímulo

Músicas de diferentes gêneros (música clássica (15); música instrumental (7); Choro, MPB, Músicas Folclóricas (7), Jazz, Blues (11)) foram pré-selecionadas por três pessoas com expertise musical para evocar ou ternura ou alegria. (Essas categorias emocionais foram escolhidas por motivos que estão fora do escopo desse trabalho.) Uma avaliação independente de 4 pessoas elegeu 20 músicas por categoria, por meio de uma pontuação de uma escala de intensidade da emoção sentida. Essas 20 músicas de 46 s de duração foram concatenadas com um *fade-in* e *fade-out* de 1 s para construir dois *medleys* a 8 min por categoria (10 músicas por *medley* com uma música inicial de 20 s).

As músicas utilizadas estão listadas nas seguintes tabelas:

Alegria 1	Nome	Artista
<i>Warmup</i>	Fugata	Piazzola
1	Corrupção	Edu Lobo
2	Bole-Bole	Ze da Velha Silverio Pontes (Choro)
3	Mercy, Mercy, Mercy	Cannonball Adderly
4	Das Wolhntemperierte Clavier-Praeludium	Bach
5	Dramophone	Caravan Palace
6	Cantaloupe Island	Herbie Hancock
7	Pret a porter de tafetá	João Bosco
8	Brandenburg Concerto No5 BWV 1050	Bach
9	A Tapdancer's Dilema	Diablo Swing Orchestra
10	Passo Rasgado	Trio Curupira

<b>Alegria 2</b>	<b>Nome</b>	<b>Artista</b>
<i>Warmup</i>	Take Five	Dave Brubeck
1	Santa Morena	Trio Madeira Brasil
2	Zauberflöte: Papagena! Weibchen!	Mozart
3	Wolverine Blues	Thomas Finn New Orleans Ensemble
4	Andre De Sapato Novo	Altamiro Carrilho
5	Freddie Freeloader	Miles Davis
6	Desvairada	Garoto
7	Fuga y misterio	Astor Piazzolla
8	Daphne	Django Reinhardt
9	There is no Business	Irving Berlin
10	Santa Tereza	Duo Foleritmia
<b>Ternura1</b>	<b>Nome</b>	<b>Artista</b>
<i>Warmup</i>	Sleeping in the field	Relaxmydog
1	The New World, Sym, No.9	Dvorak
2	Porto	Renato Braz
3	Consolation No. 3	Van Cliburn
4	Violin Concerto No1 Adagio	Bruch (Interp.Jascha Heifetz)
5	Sonata for Cello and Piano Gminor, Op.19-3	Rachmaninov
6	Sueños y mentiras	Caminho
7	Fruta Boa	Milton Nascimento
8	Concerto Per Violino No 4 Rondo	Music Therapy
9	Sonata in A Minor for Arpeggione and Piano D 821 – Allegro	Schubert
10	Ballade No. 4 Op. 52	Chopin (Rubenstein)
<b>Ternura2</b>	<b>Nome</b>	<b>Artista</b>
<i>Warmup</i>	Prelude No1 E Minor	Andres Segovia
1	Blame it on my youth	Brad Mehldau
2	2.Largo from Concerto for Oboe, Strings, Basso Continuo in D	Leonard Bernstein
3	Romance Op. 37	Jean-Pierre Rampal, Marielle Nordmann
4	Meditation from Thais	Jules Massenet
5	Albert's House	Chet Baker
6	Las 4 Estaciones Portenas IV Invierno Porteno	Bragato, Munich Piano Trio
7	Dolly Op.56-1Berceuse	Fauré
8	Pavane Pour Une Infante	Ravel
9	Palhaço	Zé Paulo Becker
10	La Fille Aux Checeux	Debussy (Alexis Weissenberg)

### 3.3 Protocolo experimental

Todos os participantes assinaram o termo de consentimento livre aprovado pelo Parecer nº 442.648 do Comitê de Ética em Pesquisa do Hospital Copa D'Or. Antes do experimento, eles receberam uma cópia digital dos *medleys* para poderem familiarizar-se com a música e treinar a tarefa de ou sentir a emoção ou de ouvir a música analisando tecnicamente para não sentir a emoção. Essas duas tarefas são descritas com mais detalhes nas instruções abaixo. Esse fator experimental da tarefa foi incluído no experimento para avaliar eventuais diferenças no processamento das músicas entre sentir ou não sentir a emoção. Junto com a cópia, eles receberam as seguintes instruções:

A participação envolverá:

#### **Treino e Critério de Inclusão:**

1 - Escutar os quatro *medleys* (dois de alegria, dois de ternura) várias vezes em casa, treinando as tarefas de

- a ) sentir a emoção
- b ) tarefa técnica: analisar a música (seguir estrutura/harmonia) para não sentir a emoção

Instruções detalhadas para as duas tarefas:

(a) Sente a emoção o mais intenso possível. Para ternura, a experiência emocional deve ser um sentimento caloroso, se sentindo próximo a familiares ou amigos (excluindo romances). Para a música de alegria, deve ser uma experiência alegre que é mais concentrada em você e que pode ser associada com (a imaginação de) movimentos juntos com a música.

(b) Tarefa técnica para não sentir a emoção: você deve se concentrar na estrutura da música, p.ex. seguir a harmonia ou a linha de base da música.

2 - Dar um retorno, se você acha que conseguiu realizar as tarefas (isto é, se as peças dos *medleys* evocaram suficientemente forte as emoções específicas e se você conseguiu se "desligar" da emoção na tarefa técnica.

#### **Caso positivo, o experimento em si, envolve:**

- 2 x 1 h algum dia antes da primeira RM: pontuar cada peça do *medley* (quanta emoção sentiu de 1 a 5, quão bem conseguiu fazer a tarefa técnica de 1 a 5)
- 4 x 1 h de RM: escutar os quatro *medleys* com uma dada tarefa na RM
- 4 x 0,5 h imediatamente depois da RM: responder questionários pós-RM
- 2 x 1 h algum dia depois das primeiras RM: pontuar novamente cada peça do *medley*

Isso dá um tempo "líquido" de experimento de 10 h por participante. É muito tempo, mas, para quem gosta de música, é um tempo bem investido!

Um dia antes da sessão de RM, os participantes pontuaram a emoção sentida e indicaram quão bem foi realizada a tarefa técnica. As músicas foram apresentadas via fones de ouvido de forma semi-automatizada com o software Presentation® (versão 14.6, www.neurobs.com). A pontuação aconteceu depois de cada música com tempo livre determinado pelo próprio participante. A ordem das músicas foi mantida consistente com os *medleys*. No apêndice, encontram-se as telas com as escalas.



*Figura 3-1: Protocolo experimental: Em cada sessão, os quatro medleys foram apresentados uma vez. A tarefa e a categoria emocional foram pseudo-aleatorizadas, garantindo que a categoria emocional alternava entre as sequências. No total das quatro sessões, o mesmo medley foi escutado por quatro vezes, sendo que duas vezes com a instrução de sentir a emoção e duas vezes com a tarefa técnica para não sentir a emoção (cor de fundo cinza).*

No total, quatro sessões na ressonância magnética foram realizadas em dias diferentes. Cada sessão incluiu quatro *medleys* diferentes, escutados durante quatro sequências funcionais (Figura 3-1). A categoria emocional alternava entre cada sequência enquanto a tarefa alternava apenas duas vezes por sessão. Após duas sessões, todas as combinações entre versões dos *medleys* e tarefas foram realizadas. As sessões três e quatro foram repetições nas quais apenas a ordem das sequências mudava. Diretamente após cada sequência, o participante pontuou verbalmente numa escala de 1 a 5 quão bem conseguiu fazer a tarefa (apenas para sequência com tarefa técnica) e quão intensamente sentia a emoção.

Após a sessão de ressonância, o participante preencheu um questionário sobre concentração, dificuldade e estratégias utilizadas nas tarefas durante a sessão realizada.

### **3.3.1 Parâmetros de aquisição de imagens**

Dados de RMf foram adquiridos num aparelho de ressonância magnética de 3T (Philips Achieva). Cada sequência incluía 245 volumes com TR = 2 s, correspondendo ao tempo total de um *medley*, TE = 22 ms, 24 cortes no plano axial com resolução espacial do plano de 3,75 × 3,75 mm e altura de 3,75 mm e 1 mm de espaçamento entre cortes. Depois de duas sequências de RMf, uma imagem anatômica com resolução de 1 mm isotrópico foi adquirida.

## **3.4 Pré-processamento RMf**

Os dados adquiridos foram exportados em formato DICOM e convertidos no formato NIfTI (<http://www.mccauslandcenter.sc.edu/mricro/mricron/dcm2nii.html>). Todos os cortes e volumes das sequências funcionais foram inspecionados visualmente com o software MIPAV para excluir possíveis artefatos. O *software* permite uma visualização completa de todos os cortes e volumes no espaço original da aquisição. No primeiro passo de pré-processamento, todos os volumes de RMf foram realinhados com o primeiro volume da primeira sessão de cada sujeito. Depois, tendências lentas e frequências baixas foram removidas com um filtro de Savitzky-Golay, conforme aplicado em (ÇUKUR et al., 2013). Efeitos fisiológicos no sinal BOLD foram removidos conforme descrito abaixo. Da mesma forma, o sinal global foi subtraído. Por último, filtros de suavização espacial e temporal foram aplicados aos dados. A análise principal foi aplicada tanto no espaço original do sujeito como também depois da

normalização linear para o espaço MNI com o método *mcflirt* do pacote FSL (JENKINSON et al., 2002; SMITH et al., 2004)). Todos os passos são descritos com maiores detalhes nos parágrafos seguintes.

### 3.4.1 Inspeção visual do dado bruto

A inspeção visual é um passo importante para garantir a qualidade dos dados. Durante a inspeção, artefatos técnicos, p.ex. *spikes*, ou movimentos abruptos entre um e outro volume são facilmente identificados.

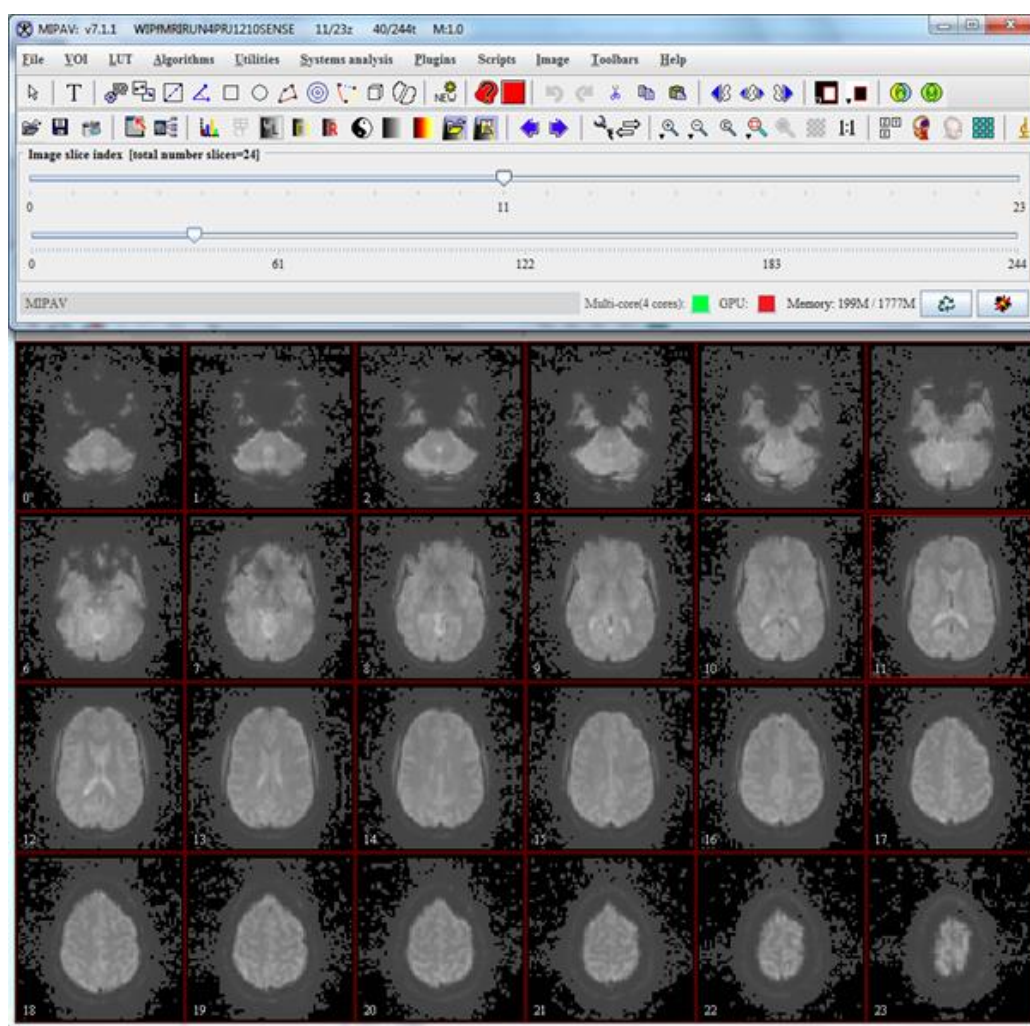


Figura 3-2: Visualização do dado bruto no espaço original da aquisição: 24 cortes estão agrupados em formato mosaico. A barra superior mostra o volume atual (40) e permite uma passagem por todos os volumes de forma contínua.

### 3.4.2 *Realinhamento*

O primeiro volume da primeira sessão foi utilizado como referência para, conseqüentemente, realinhar todos os volumes restantes. Dessa forma, a localização do *voxel* é mantida na mesma posição, que poderia sofrer alteração por eventuais movimentos durante e entre as sessões. Esse procedimento também garante a correspondência entre posições de *voxels* entre diferentes sessões e dias. O algoritmo *mcflirt* (JENKINSON et al., 2002) foi utilizado desta maneira:

```
mcflirt -in volumes4D.nii -out volumes4D_mc.nii -reffile volume1.nii -mats -
plots
```

### 3.4.3 *Remoção de tendências/frequências baixas*

Tendências lentas foram removidas de cada *voxel* separadamente com um filtro de Savitzky-Golay, usando 121 pontos no tempo (correspondendo a 242 s) e um polinômio ordem 3 (ÇUKUR et al., 2013).

### 3.4.4 *Correção por efeitos fisiológicos*

A correção de efeitos fisiológicos no sinal BOLD é baseada em sinais periféricos (respiração, ECG, GSR) coletados simultaneamente com RMf. Para isto, cada *voxel* é modelado por um conjunto de regressores, que representam diferentes fontes de ruído fisiológico no sinal BOLD. Alguns desses regressores são estabelecidos seguindo a abordagem RETROICOR (*inglês*: RETROspective Image CORrection) (GLOVER; LI; RESS, 2000), que foca na periodicidade e fase dos sinais. Primeiramente, a fase de cada ciclo dos sinais de respiração e ECG é extraída e, em seguida, expandida por séries de Fourier de baixa ordem (quarta e terceira ordem, respectivamente). A fase do ECG  $\varphi_c$  no momento  $t$  é determinada pelo pico R anterior  $t_1$  e  $t_2$ , seguindo:

$$\varphi_c(t) = 2\pi \frac{t - t_1}{t_2 - t_1} \quad (2)$$

A fase da respiração  $\varphi_r$  é normalizada pelo histograma  $H$  sobre as amplitudes do sinal:

$$\varphi_r(t) = \text{sgn}(dR/dt) \cdot \pi \frac{\int_{R_{\min}}^{R(t)} H(R) dR}{\int_{R_{\min}}^{R_{\max}} H(R) dR} \quad (3)$$

Por último, para cada sinal, a série Fourier é expandida para que frequências mais altas, possivelmente introduzidas por efeitos de *aliasing*, possam ser capturadas:

$$x(t) = \sum_{m=1}^p A_m \cdot \cos(m\varphi) + B_m \cdot \sin(m\varphi) \quad (4)$$

Adicionalmente a interação cardiorrespiratória foi modelada com uma interação de primeira ordem:

$$x_{c.xr}(t) = A \cdot \cos(\varphi_c) \cos(\varphi_r) + B \cdot \sin(\varphi_c) \cos(\varphi_r) + C \cdot \cos(\varphi_c) \sin(\varphi_r) + D \cdot \sin(\varphi_c) \sin(\varphi_r) \quad (5)$$

Essas etapas foram realizadas com a ferramenta PhysIO (KASPER et al., 2009), que é parte do pacote TAPAS, o qual dispõe código aberto em Matlab® (<http://www.translationalneuromodeling.org/tapas/>). Além da abordagem RETROICOR, a ferramenta implementa a convolução dos sinais respiratório e cardíaco com funções de respostas específicas. A importância da função da resposta respiratória (RRF) foi investigada em (BIRN et al., 2006), e o formato determinado em (BIRN et al., 2008) como:

$$RRF(t) = 0.6 \cdot t^{2.1} \cdot e^{-t/1.6} - 0.0023 \cdot t^{3.54} \cdot e^{-t/4.25} \quad (6)$$

Os sinais extraídos foram inspecionados visualmente, e eventuais problemas (p.ex. detecção de pico R), corrigidos.

O sinal da resposta galvânica foi exportado e convoluído com a função da resposta hemodinâmica do SPM. No total, 21 (RETROICOR: 18; RRF: 1; CRF: 1; GSR: 1) regressores foram utilizados para subtrair efeitos fisiológicos de não interesse. Primeiro, para cada *voxel* a contribuição  $\beta$  de cada regressor em  $X$  foi estimada com o



método dos mínimos quadrados. Depois, essa contribuição foi subtraída para obter um sinal corrigido  $Y_c$ :

$$\begin{aligned} Y &= X \cdot \beta + \varepsilon \\ Y_c &= Y - X \cdot \beta \end{aligned} \quad (7)$$

### **3.4.5 Correção do sinal global (cérebro inteiro)**

O sinal global foi subtraído por uma regressão simples por *voxel* (FOX et al., 2009). A média de todos os *voxels* do cérebro inteiro foi calculada ao longo do tempo e utilizado como único regressor para subtrair a contribuição para cada *voxel* separadamente, seguindo a equação (6).

### **3.4.6 Suavização espacial**

Um filtro espacial pode aumentar a razão sinal-ruído. Por isso, um filtro gaussiano com 8 mm de FWHM foi aplicado ao dado, utilizando o seguinte comando:

```
fslmaths input.nii -kernel gauss 8/2.3548 -fmean output.nii
```

### **3.4.7 Suavização temporal**

O sinal BOLD tem uma evolução lenta, chegando ao pico uns 4-6 s depois de um estímulo. Por isso, uma suavização temporal pode aumentar eventualmente a razão sinal-ruído (ALLURI et al., 2013). Foi aplicado um filtro gaussiano no tempo de 5 s FWHM:

```
fslmaths input.nii -bptf -1 5/2 output.nii
```

### **3.4.8 Normalização linear para espaço MNI**

A normalização para o espaço MNI ajuda para comparar resultados com atlas estabelecidos e entre sujeitos. Para minimizar a distorção dos *voxels* originais de aquisição, uma normalização linear foi aplicada da seguinte forma:

```
flirt -in refile.nii -ref mni_template.nii -omat subj2mni
```

```
flirt -in filein.nii -ref mni_template.nii -applyxfm -init subj2mni -out fileout.nii
```

### **3.5 Codificação de características musicais**

Os modelos de codificação requerem uma descrição quantitativa dos estímulos ao longo do tempo. Dessa forma, é possível estimar a relação entre estímulo e resposta cerebral (confere Seção 1.2). A seguir, todos os passos desse procedimento serão descritos em detalhe.

#### **3.5.1 Extração de características acústicas das músicas**

Dos quatro *medleys*, foram extraídas diversas características acústicas para obter uma descrição ampla do estímulo. O procedimento seguiu basicamente os estudos de (ALLURI et al., 2012, 2013), nos quais foram extraídas características timbrais, rítmicas e tonais. Enquanto as características timbrais são calculadas sobre janelas de curta duração de 25 ms e uma sobreposição de 50%, as características rítmicas e tonais utilizam janelas de 3 s. Essas durações são estabelecidas na área da extração de características musicais. A Tabela 3-1 descreve todas as características extraídas com mais detalhes.

Tabela 3-1 – Descrição das características musicais, seguindo *ALLURI et al. (2012)*.

<b>NOME</b>	<b>DESCRIÇÃO</b>	<b>JANELA</b>
<i>Zero-Crossing-Rate</i>	Quantidade das vezes que o sinal de áudio passa pelo zero por unidade de tempo.	25 ms
<i>Spectral centroid</i>	Centro de massa da densidade espectral.	25 ms
<i>Brightness</i>	Proporção entre energia em frequências maiores e menores de 1500 Hz.	25 ms
<i>Spectral Entropy</i>	Entropia da densidade espectral, utilizando a entropia relativa de Shannon.	25 ms
<i>Spectral Rolloff</i>	O valor da frequência embaixo do qual tem 85% de energia na densidade espectral. É uma forma de descrever a forma da densidade espectral (TZANETAKIS; COOK, 2002).	25 ms
<i>Spectral Flux</i>	Medida de alteração da densidade espectral, definida pela diferença quadrada entre densidades espectrais de duas janelas sucessivas.	25 ms
<i>Spectral Spread</i>	Desvio padrão da densidade espectral.	25 ms
<i>Spectral Flatness</i>	Entropia de Wiener, definida pela razão entre média geométrica e aritmética. O valor máximo é obtido para ruído branco, e o valor mínimo para um tom de uma única frequência.	25 ms
<i>Sub-band flux (1 para cada banda)</i>	Medida de alteração da densidade espectral, calculada separadamente para 10 bandas de frequência. A percepção das bandas com frequências baixas de 50Hz-200Hz foi descrita como “ <i>Fullness</i> ”, e nas frequências mais altas de 1600 Hz – 6400 Hz como “ <i>Activity</i> ” (ALLURI; TOIVIAINEN, 2010)	25 ms
<i>Key clarity</i>	Medida da clareza tonal, baseada em estimativas de alturas de tons.	3 s
<i>Pulse clarity</i>	Medida da clareza do ritmo, baseada na autocorrelação do envelope.	3 s

Todas as características acústicas foram extraídas com a ferramenta MIRtoolbox 1.4 (LARTILLOT; TOIVIAINEN, 2007; LARTILLOT; TOIVIAINEN; EEROLA, 2008), disponível no ambiente Matlab®.

Após a extração, as características foram convoluídas com uma função de resposta hemodinâmica com formato de duas funções gama (em inglês, conhecida como *double-gamma HRF*) e, finalmente, reamostradas para 0,5 Hz para corresponder ao sinal BOLD que foi adquirido com um TR de 2 s.

### 3.5.2 Regressão linear múltipla

Três diferentes estimadores foram avaliados para a regressão linear múltipla. O primeiro e mais básico é o estimador que minimiza a soma do erro quadrado. O segundo minimiza a norma L2 (regressão *ridge*), e o último minimiza a norma L1 e L2 (*Glmnet*) (HASTIE; TIBSHIRANI; FRIEDMAN, 2009):

1. Mínimo dos erros quadrados:

$$\arg \min_{\beta} \left( \sum_{i=1}^N (Y_i - X_i \beta)^2 \right) \quad (8)$$

2. Regressão *ridge* (norma L2)

$$\arg \min_{\beta} \left( \sum_{i=1}^N (Y_i - X_i \beta)^2 + \lambda \|\beta\|_2^2 \right) \quad (9)$$

3. Regressão por *Glmnet* (minimiza norma L1 e L2):

$$\arg \min_{\beta} \left( \sum_{i=1}^N (Y_i - X_i \beta)^2 + \lambda \left[ (1 - \alpha) \|\beta\|_2^2 / 2 + \alpha \|\beta\|_1 \right] \right) \quad (10)$$

Enquanto a norma L2 tende a reduzir os coeficientes de forma geral, a norma L1 (*lasso*) tem a tendência de zerar alguns coeficientes. A regressão por *Glmnet* usa uma combinação das duas penalizações. Vale destacar que a regressão *ridge* e *Glmnet* são métodos que evitam um *overfitting* quando o número de parâmetros no modelo é grande em comparação ao número de pontos (GREEN, 1991).

### **3.5.3 Validação cruzada**

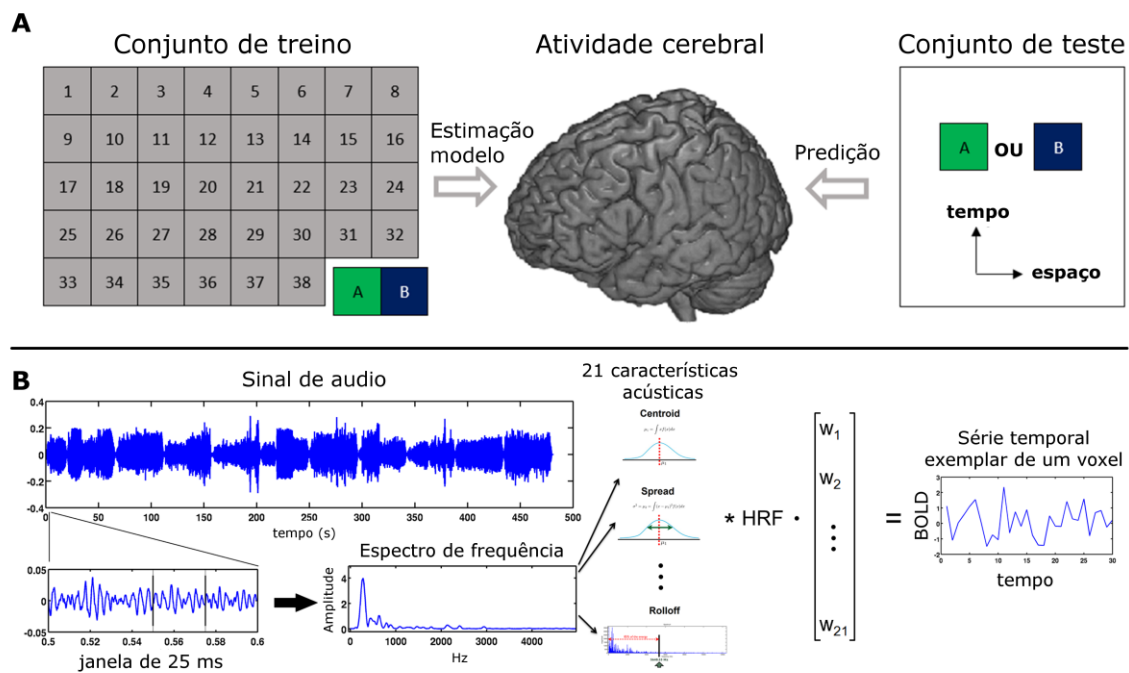
A validação cruzada é uma técnica muito conhecida na área de aprendizagem de máquina. Ela é crucial para determinar a generalização do modelo. Para este fim, os dados são divididos em duas partes: Uma parte é reservada para o treino, ou seja, a estimação do modelo, e a outra parte é utilizada para validar o modelo estimado. As divisões dos dados são exclusivas e, geralmente, se usa alguma estratégia de repetição. Por exemplo, uma divisão de 80% para estimação do modelo e os restantes 20% para validação do modelo pode ser repetido em 5 vezes, garantindo que o dado da validação é sempre diferente. Cada validação por si é uma generalização e, no caso da regressão linear múltipla, ela é expressa em termos de correlação da predição: a partir do modelo estimado durante o treino, se prediz o dado da validação.

### **3.5.4 Metodologia “Maestro qual é a música?”**

A validação cruzada foi organizada de forma que duas músicas ficaram fora dos dados de treinamento. Nestes dados de treinamento, foram estimados os coeficientes da regressão linear múltipla e, conseqüentemente, aplicados para identificar qual das duas músicas era mais provável de evocar a resposta cerebral observada. No total, 780 ( $40 \times 39 / 2$ ) pares foram testados. A regressão linear múltipla foi estimada pelo método do mínimo erro quadrático. Não houve necessidade de utilizar uma técnica de regularização, já que o número de pontos (4 repetições x 38 músicas x 23 pontos = 3.496 pontos) era ordens de grandezas maior do que o número de variáveis (21). Durante o treino, todos os dados de todas as sessões foram agregados sem considerar os fatores experimentais. A ordem da seleção dos *voxels* utilizados durante a fase de decodificação foi estabelecida durante o treino, utilizando cinco divisões com proporções de 80% (2.796 pontos) e 20% (699 pontos) dentro do dado treino. A média das cinco predições (correlação temporal entre os 699 pontos preditos e observados) foi calculada para determinar a ordem final dos *voxels*.

Na codificação, para cada par de músicas A e B, o algoritmo determinava se (1) a atividade cerebral durante a música A foi melhor predita pelas características acústicas da música A do que da música B e se (2) a atividade cerebral durante a música B foi melhor predita pelas características acústicas da música B do que da música A. Durante a identificação, o número de pontos no tempo e número de voxels utilizados para a predição, foram progressivamente aumentados. Essa abordagem sistemática permite investigar o desempenho do modelo em função do tempo e espaço. Uma música é identificada corretamente quando ela é melhor predita pelas características correspondentes do que pelas características de uma outra música. A decisão foi baseada na correlação espaço-temporal entre a predição e o dado de RMf observado.

A maioria das características foram derivadas do espectro de frequência utilizando janelas de 25 ms, posteriormente convoluídas com a resposta hemodinâmica funcional (HRF) e reamostradas para 0,5 Hz, taxa de amostragem utilizada nas imagens de ressonância funcional. Este procedimento é ilustrado na Figura 3-3.



*Figura 3-3: Modelo de codificação e decodificação (A) Ilustração do procedimento para identificar as músicas. Em cada uma das 780 iterações o modelo foi estimado para cada voxel independentemente com 38 das 40 músicas (Conjunto de treino). Baseado no modelo estimado, a atividade cerebral para as duas músicas de teste foi predita. (B) Modelo de codificação: Do sinal de áudio das 40 músicas foram extraídas 21 características acústicas. O conjunto das 21 características acústicas foi utilizado numa regressão linear múltipla para modelar a atividade cerebral (sinal BOLD).*

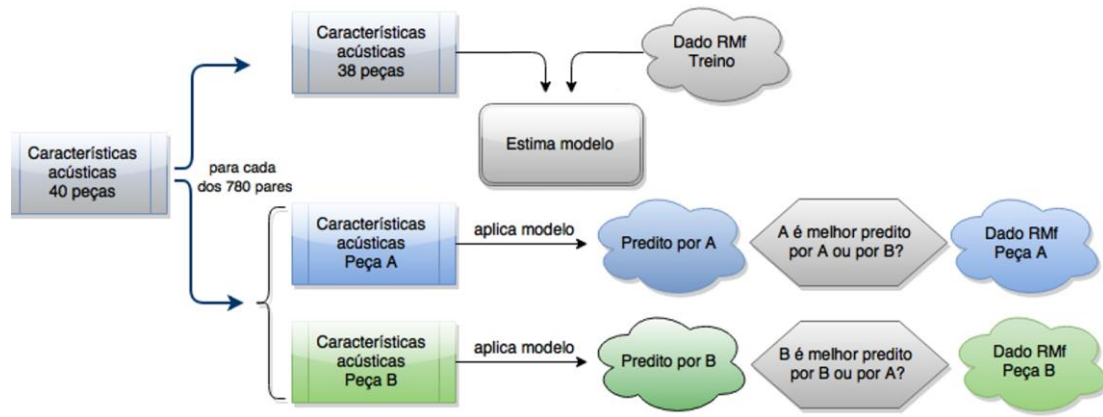


Figura 3-4: Identificação de músicas: Em cada uma das 780 iterações o modelo é estimado com 38 das 40 músicas (dado de treino). O modelo estimado é aplicado para prever a atividade cerebral para as duas músicas de teste. A música 'A' é identificada corretamente se a predição pelas características correspondentes de 'A' é melhor, isto é, tem correlação maior com o dado de RMf, do que pelas características acústicas da música não correspondentes de 'B'.

A decisão sobre a predição melhor de uma música **A** foi baseada na correlação entre o sinal observado  $y_a$  durante a música **A** e o sinal predito pelas características dessa música  $\hat{y}_a$  em comparação com o sinal predito  $\hat{y}_b$  pelas características da outra música **B**:

$$\text{correto} = 1, \text{ se } \text{corr}(y_{a_{\text{voxel,tempo}}}, \hat{y}_{a_{\text{voxel,tempo}}}) > \text{corr}(y_{a_{\text{voxel,tempo}}}, \hat{y}_{b_{\text{voxel,tempo}}}) \quad \text{EQ 1}$$

Nesse caso, os sinais de múltiplos *voxels* e pontos no tempo foram reformatados de uma matriz para um vetor com apenas uma dimensão antes de calcular a correlação de Pearson.

A acurácia da identificação foi determinada pela fração de músicas corretamente identificadas sobre o número total de identificações:

$$\text{acurácia} = \frac{\#(\text{identificado corretamente})}{40 * 39} \quad \text{EQ 2}$$



Para investigar a dependência da extensão da área cerebral no córtex auditivo e tempo necessário para a identificação, os cálculos da decisão e da acurácia foram aplicados separadamente, variando o número de *voxels* na faixa de 2 a 300 e adicionando ponto a ponto no tempo de 2 s a 46 s (correspondendo a volume 1 a 23). Os *voxels* foram selecionados na ordem decrescente das predições obtidas numa validação cruzada intra-treino como descrito na seção anterior.

## 4 Resultados

### 4.1 Escalas comportamentais

Até no mínimo um dia antes da sessão na RM, os participantes responderam para cada uma das 20 músicas por categoria quão bem conseguiram sentir a emoção e realizar a tarefa técnica para não sentir a emoção. A Figura 4-1 mostra que, na média, os sujeitos conseguiram realizar as tarefas para cada música. Os sujeitos pontuaram em tempo livre depois de cada peça numa escala de 1 a 5 quão bem conseguiram realizar a tarefa. Na média, os sujeitos conseguiram realizar as tarefas igualmente bem para todas as peças. Apenas o sujeito 5, indicado pela anotação no gráfico, teve uma emoção menos intensa para algumas peças.

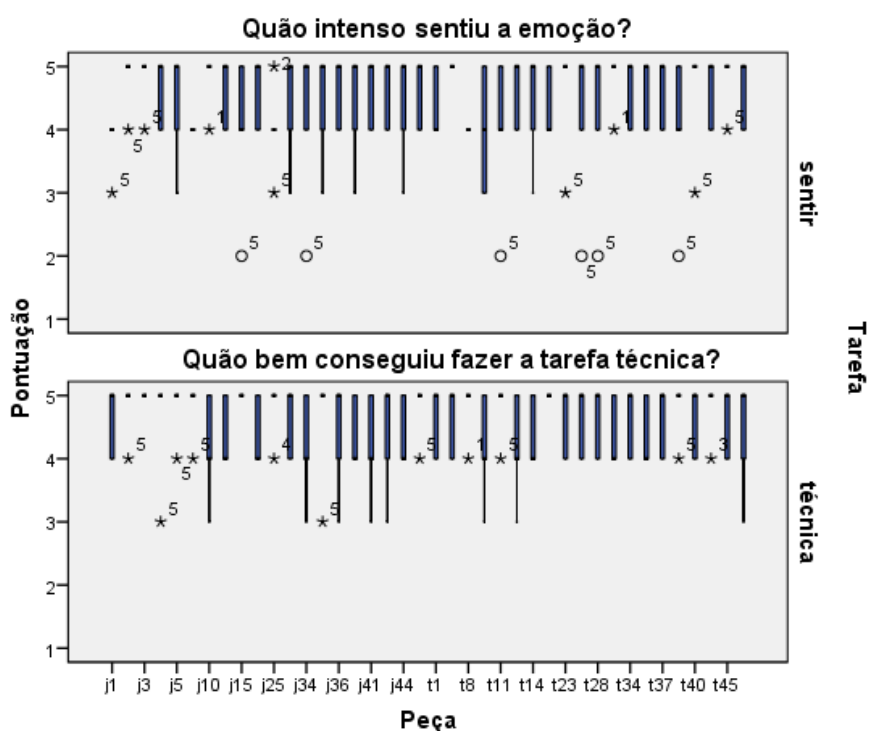
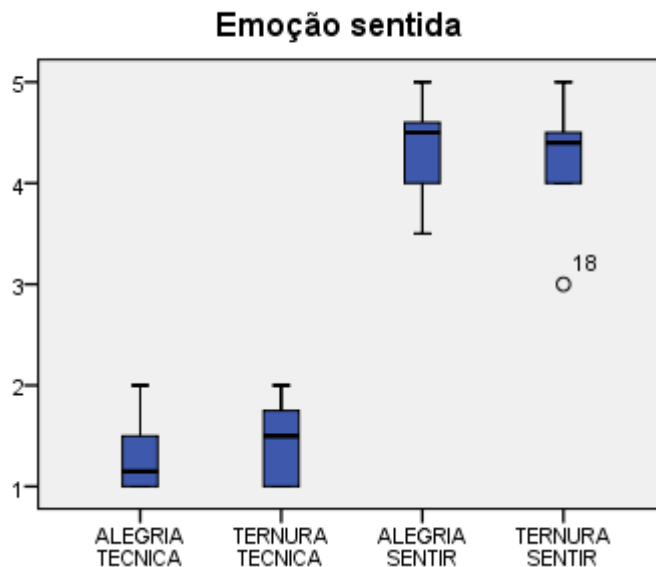


Figura 4-1: Box-plots mostrando pontuações individuais antes da RM para as 20 peças de alegria e 20 peças de ternura, indicadas pela letra 'j' e 't'. A barra mais larga mostra a região entre o primeiro e o terceiro quartis; outliers são marcados com círculos; outliers extremos (três vezes maior do que o interquartil) com asterisco – o número identifica o sujeito.

Após cada sequência na RM, os sujeitos relataram, numa escala de 1 a 5, a emoção sentida. A *Figura 4-2* mostra que os sujeitos conseguiram não sentir emoção na condição da tarefa técnica e sim na tarefa de sentir.



*Figura 4-2: Box-plot mostrando a intensidade da emoção sentida de uma escala de 1 a 5. Cinco sujeitos são agregados, ou seja, cada barra resume 20 valores (4 sessões por sujeito e por condição) – um ponto (“18” correspondendo ao sujeito 5) é marcado como outlier. Os sujeitos foram instruídos sobre a tarefa antes de cada sequência e responderam imediatamente após “quão intenso sentiram a emoção”. As respostas mostram que as tarefas “TECNICA” e “SENTIR” foram realizadas com sucesso.*

Depois da sessão RM, os sujeitos pontuaram diferentes dimensões emocionais ao ouvir as músicas. Os resultados da *Figura 4-3* mostram que as músicas escolhidas para evocar alegria ou ternura tem um perfil bem característico marcado pela categoria alvo.

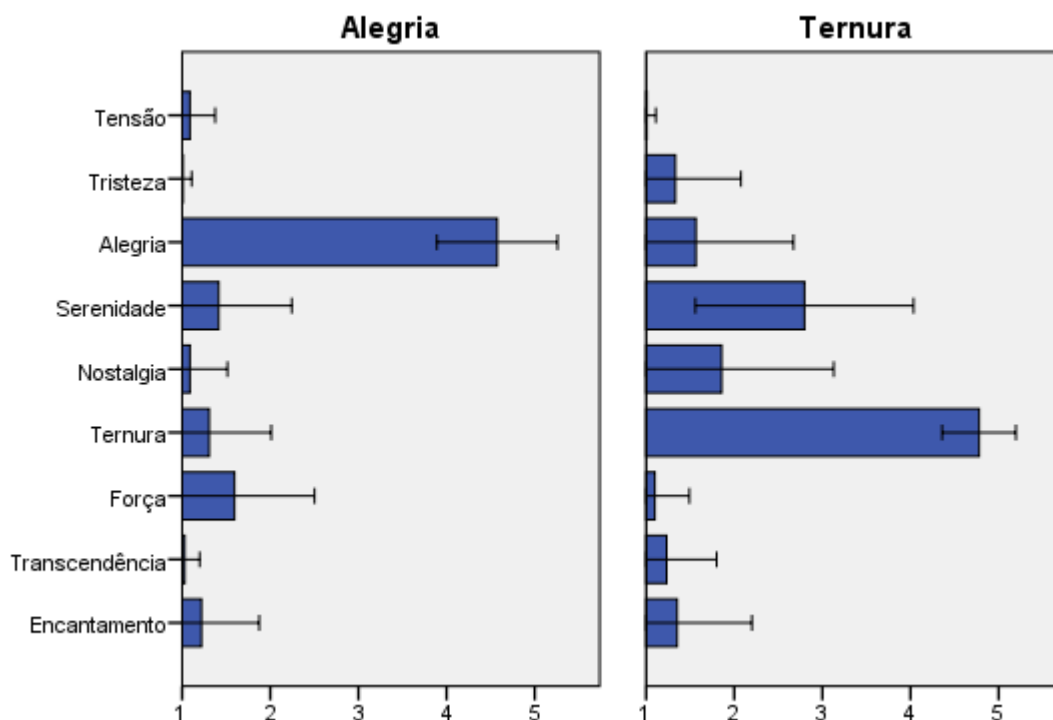


Figura 4-3: Pontuações da escala GEMS-9 descrevendo diferentes dimensões das emoções sentidas. A barra mostra a média sobre 100 valores (5 sujeitos e 20 músicas por categoria) e o desvio padrão. As músicas de alegria recebem maior pontuação na dimensão alegria, seguido por força. Na ternura a serenidade e nostalgia receberam maiores valores depois da dimensão alvo de ternura.

## 4.2 Respostas fisiológicas durante RMf

As respostas fisiológicas foram analisadas separados por condição e tarefa. Cada sequência de 8 min foi subdividida em janelas de 15 s. Dentro de cada janela, os tempos entre picos R do ECG (RR), o desvio padrão dos tempos RR e a frequência com maior energia da respiração foram extraídos. As médias por sequência dessas métricas entraram na análise estatística. Uma ANOVA de medidas repetidas com os fatores emoção, tarefa e sequência indicou uma interação significativa entre emoção e tarefa ( $F(1,5) = 19,17$ ;  $p = 0,007$ ) e um efeito principal da emoção ( $F(1,5) = 21,34$ ;  $p = 0,006$ ) para a frequência respiratória. O fator tarefa apenas mostrou uma tendência ( $F(1,5) = 5,16$ ;  $p = 0,07$ ). A comparação das médias das emoções mostrou uma respiração mais lenta para ternura do que para alegria (ternura:  $M = 0,30$ ; IC 95% =

$[0,26; 0,34]$ ; *alegria*:  $M = 0,33$ ;  $IC\ 95\% = [0,30; 0,36]$ ) e uma diminuição ainda maior pelo efeito da tarefa na condição de ternura (*ternura sentir*:  $M = 0,28$ ;  $IC\ 95\% = [0,23; 0,33]$ ; *ternura técnica*:  $M = 0,33$ ;  $IC\ 95\% = [0,29; 0,37]$ ).

Nenhuma interação e nenhum efeito principal foram indicados para os tempos entre picos RR. Já para o desvio padrão entre picos RR, efeitos significativos foram obtidos para a tarefa ( $F(1,5) = 6,53$ ;  $p = 0,051$ ) e a interação entre emoção e tarefa ( $F(1,5) = 10,36$ ;  $p = 0,02$ ). A comparação das médias mostrou variação maior dos picos RR para a condição de sentir ( $M = 0,043$ ;  $IC\ 95\% = [0,033; 0,052]$ ) e um efeito maior da tarefa na condição de ternura (*ternura sentir*:  $M = 0,046$ ;  $IC\ 95\% = [0,036; 0,055]$ ; *ternura técnica*:  $M = 0,037$ ;  $IC\ 95\% = [0,028; 0,046]$ ).

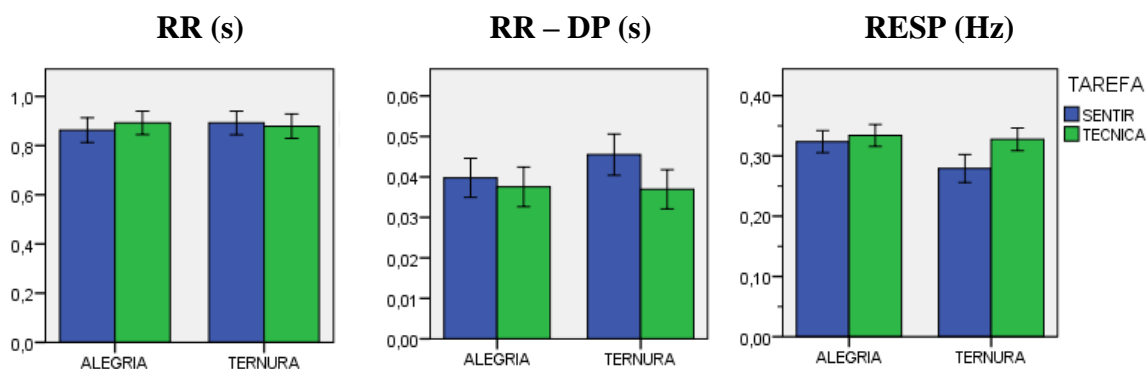


Figura 4-4: Respostas fisiológicas. Barras mostrando média e linhas pretas os intervalos de confiança (95%) entre sujeitos. As métricas de esquerda para direita são: RR = tempos entre picos R do ECG em segundos; RR - DP: Desvio padrão de tempos entre picos RR; RESP = Frequência respiratória com maior energia em Hz.

### 4.3 Comparação de modelos

Nesta seção, as primeiras avaliações do dado são apresentadas. Essas investigações foram realizadas com o objetivo de ganhar uma primeira intuição da relação do modelo e do dado. Por razões cronológicas do estudo, as análises incluem apenas cinco sujeitos. Dois tópicos são relacionados à modelagem da codificação descrita na seção 3.5. Primeiro, o impacto de alguns passos do pré-processamento é apresentado. Depois, a consistência do sinal BOLD é colocada como referência para os modelos de codificação. Em seguida, diferentes modelos são comparados.

### **4.3.1 Comparação: correlações BOLD-BOLD e codificação**

A modelagem da codificação de características acústicas tem como objetivo prever a atividade BOLD a partir da evolução temporal da música. Essa previsão é expressa, naturalmente, em correlações de previsões de séries temporais. Para determinar a qualidade da previsão de forma geral, precisa-se ter uma referência, já que a modelagem pode ser imperfeita por vários motivos: primeiro, por se tratar de um conjunto finito de variáveis que descrevem as características acústicas, é provável que este conjunto não descreva por completo as músicas; segundo, a modelagem do formato da resposta hemodinâmica – mesmo que muito estabelecido – pode introduzir alguma imperfeição; terceiro, a percepção do sujeito em diferentes momentos (dias) pode sofrer alteração e, subsequentemente, modificar a resposta cerebral; e, por último, o próprio sinal BOLD é sujeito a várias fontes de ruídos técnicos e biológicos. Os primeiros dois pontos são relacionados à modelagem matemática e à descrição dos estímulos, enquanto os últimos dois pontos, à variável dependente, o sinal BOLD. Para julgar sobre os primeiros dois pontos, deve-se ter, então, primeiro, uma referência da consistência da variável dependente. Esta referência existe neste experimento pelo fato de que o mesmo protocolo experimental foi repetido uma vez. Desta forma, a correlação entre o sinal BOLD entre as repetições reflete a consistência da variável dependente. Neste caso, a correlação BOLD a BOLD sobre 460 pontos no tempo (23 pontos por música, 20 músicas) foi calculada, separadamente, por categoria emocional (ternura, alegria) e tarefa experimental (sentir, tarefa técnica) e, conseqüentemente, comparada com a correlação de previsão obtida pela modelagem de codificação (*Figura 4-5*). O modelo de codificação foi estimado pelo método dos mínimos quadrados, depois de ter confirmado que os métodos de penalização *ridge* e *glmnet* não obtiveram resultados melhores.

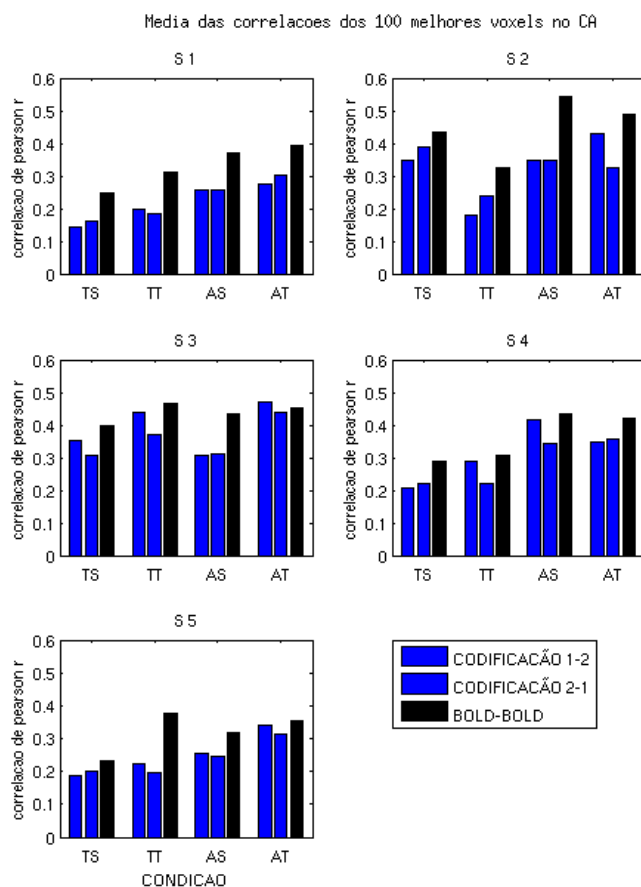


Figura 4-5: Comparação da correlação de predição da codificação (em azul) com a correlação BOLD a BOLD (em preto) para os sujeitos S1 a S5, separado pelas condições de ternura sentir (TS), ternura com tarefa técnica (TT), alegria sentir (AS) e alegria tarefa técnica (AT). As barras indicam a média dos 100 melhores voxels dentro de uma região ampla ao redor do córtex auditivo (CA), incluindo giro temporal superior, médio e giro de Heschl. O modelo foi estimado utilizando a primeira rodada do experimento e validado na repetição (1-2) e na direção inversa (2-1). A correlação BOLD-BOLD reflete a consistência da variável dependente na modelagem de codificação e serve como referência. A diferença entre a codificação e BOLD-BOLD pode ser explicado pela incompletude da descrição do estímulo, as características acústicas, e pela imperfeição do modelo da resposta hemodinâmica.

**Discussão parcial:**

A partir da *Figura 4-5*, podemos resumir alguns pontos:

- a) Na maioria das vezes, o modelo da codificação está inferior à correlação BOLD-BOLD. Existem exceções como p.ex. na condição alegria com a tarefa técnica nos sujeitos S3 e S5, nos quais a codificação tem resultados parecidos ou até melhores do que a correlação BOLD-BOLD.
- b) A consistência do sinal BOLD varia entre sujeitos.
- c) A consistência do sinal BOLD varia entre categorias emocionais e tarefas, porém não sistematicamente.
- d) Existe alguma variabilidade intrínseca nas correlações de predição entre 1-2 e 2-1, porém, de forma geral, a codificação acompanha as correlações BOLD-BOLD. Isto é um resultado importante, porque mostra que o modelo não tem qualquer viés entre categorias emocionais ou entre tarefas experimentais.

**4.3.2 Comparação de modelos de codificação**

A questão da incompletude e imperfeição da modelagem da resposta hemodinâmica foi investigada explicitamente, comparando a modelagem original com duas modificações específicas:

- a) **Extensão da descrição do estímulo:** aumento de número de variáveis descrevendo outras características acústicas, como: 1) inclinação do ataque, 2) tempo inicial de notas, 3) densidade de notas, 4) tempo, 5) novidade estrutural entre segmentos de 1 s, 6) novidade estrutural entre segmentos de 5 s, 7) mudanças de harmonia.
- b) **Modelo FIR:** modelagem da resposta hemodinâmica com uma resposta ao impulso finito com atraso de 4 s, 6 s e 8 s, seguindo (ÇUKUR et al., 2013).



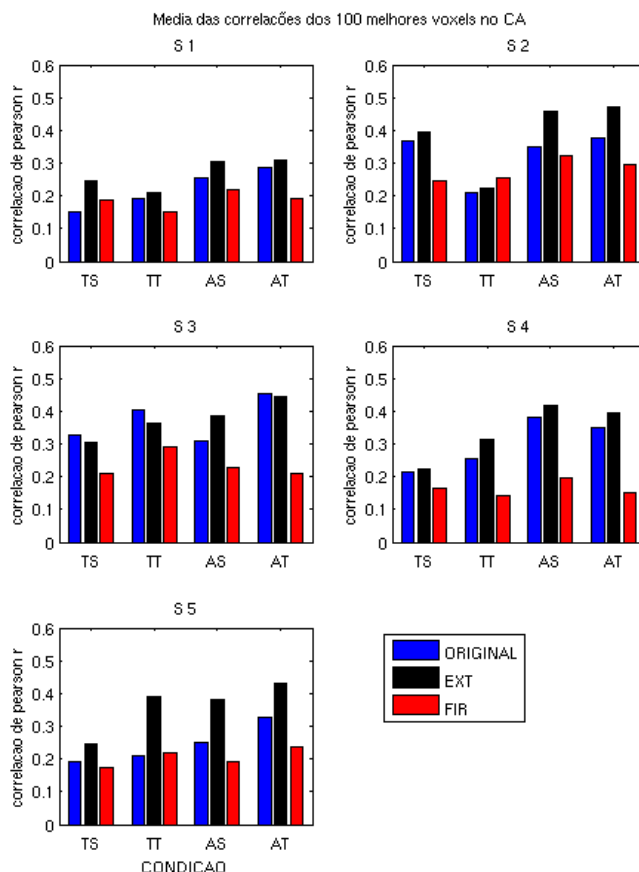


Figura 4-6: Comparação do modelo original (azul) com um conjunto de variáveis acústicas mais extenso (EXT em preto) e o modelo com resposta ao impulso finito (FIR em vermelho). O modelo com o conjunto mais extenso é superior aos outros modelos em todos os sujeitos exceto S3.

A Figura 4-6 mostra a comparação do modelo original com o modelo com um conjunto de variáveis mais extenso e o modelo com resposta ao impulso finito. Enquanto o modelo EXT obtém predições melhores para todas as condições em todos os sujeitos exceto o sujeito S3, o modelo FIR é inferior em todos os casos a não ser para condição TS em S1 e TT em S2.

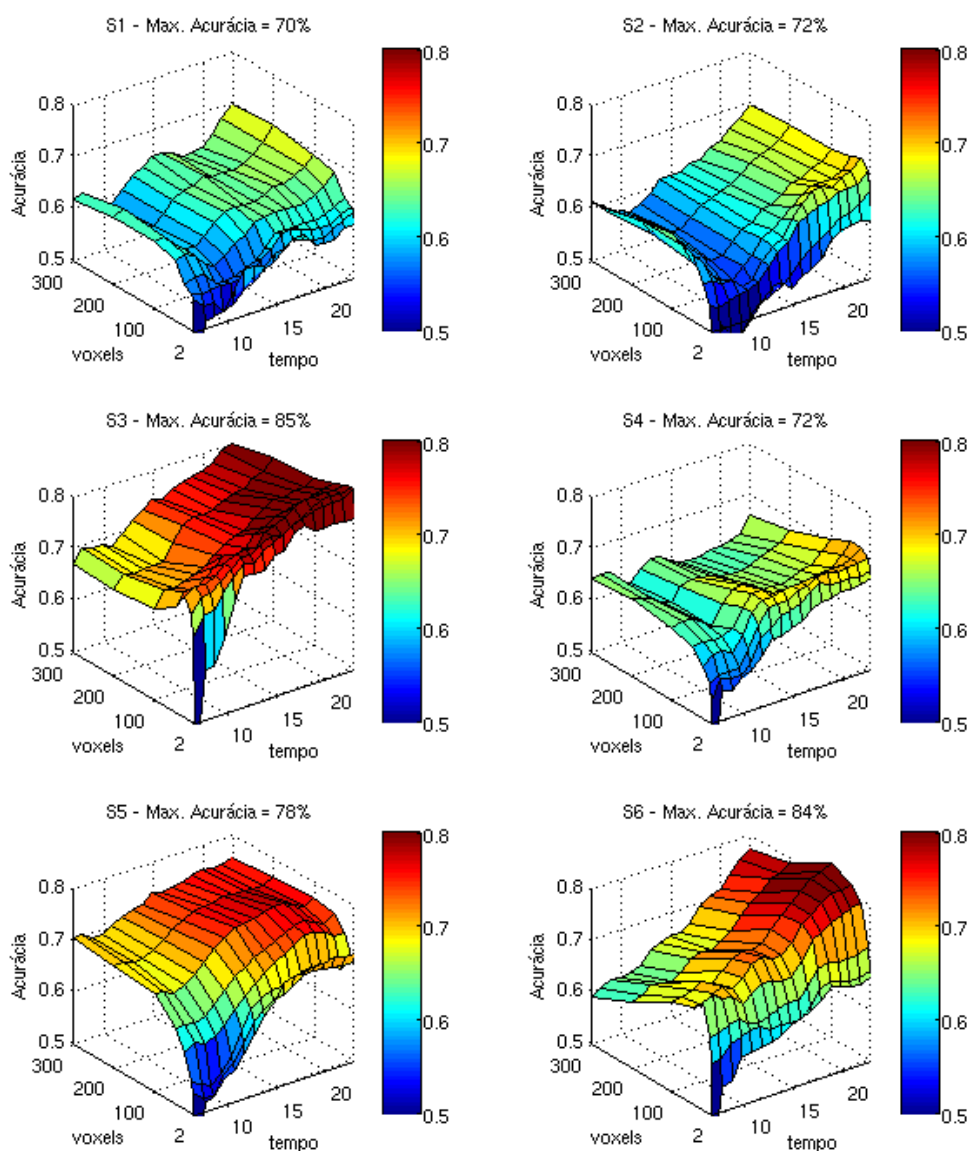
#### Discussão parcial:

Os resultados mostraram que o modelo mais extenso é o melhor para a predição da resposta cerebral. Isso sugere que as variáveis acústicas adicionadas contêm uma informação relevante que não existiu ainda no conjunto original. Em estudos futuros, poder-se-ia investigar mais a fundo quais outras características acústicas poderiam melhorar ainda mais o resultado.

Em termos da modelagem do formato da resposta hemodinâmica, pode se concluir que o modelo original com a convolução com as duas funções gama (como descrito na seção 3.5.1) é superior à modelagem com a resposta ao impulso finito (FIR). Talvez esse resultado possa ser explicado pelo fato de que, no nosso experimento, as variáveis do modelo são contínuas, enquanto a modelagem FIR no estudo de ÇUKUR et al. (2013) utiliza variáveis binárias. Além disso, nosso pré-processamento utilizou um filtro de suavização temporal que, em teoria, deve favorecer a convolução contínua.

#### **4.4 Aplicação: “Maestro, qual é a música?”**

Os resultados na *Figura 4-7* mostram que a acurácia inicial, utilizando apenas um ponto no tempo e dois *voxels*, está próximo do nível de chance (50% para a identificação binária). Adicionando mais *voxels* e mais pontos no tempo, a acurácia aumenta e atinge níveis bem acima da significância estatística de 62,5% (estabelecido com testes de permutações,  $\alpha = 0,05$ ) para 300 *voxels* e 23 pontos no tempo. Pode-se, ainda, observar que, no caso de se utilizar apenas um ponto no tempo, a acurácia, necessariamente, depende da dimensão espacial, e sobe substancialmente de 2 a 100 *voxels* e continua praticamente no mesmo nível a partir daí. A *Figura 17* resume os resultados dos seis sujeitos e mostra o desempenho do modelo em função do tempo e espaço separadamente. Fica evidente a relação linear entre acurácia e tempo. Espacialmente, o modelo melhora de forma substancial no início e satura depois de aproximadamente 100 *voxels*.



*Figura 4-7: Acurácias de identificação das músicas para os sujeitos S1 a S6. A acurácia é perto de 50% para um ponto no tempo e dois voxels. Depois, ela aumenta com o número de voxels (eixo esquerdo) e o número de pontos no tempo (eixo direito). Os números de voxels são 2, 10, 25, 50, 100, 200 e 300. O eixo do tempo começa no volume 7 (os primeiros 6 volumes são descartados para excluir a transição entre músicas, conforme explicado na discussão) e termina no volume 23.*

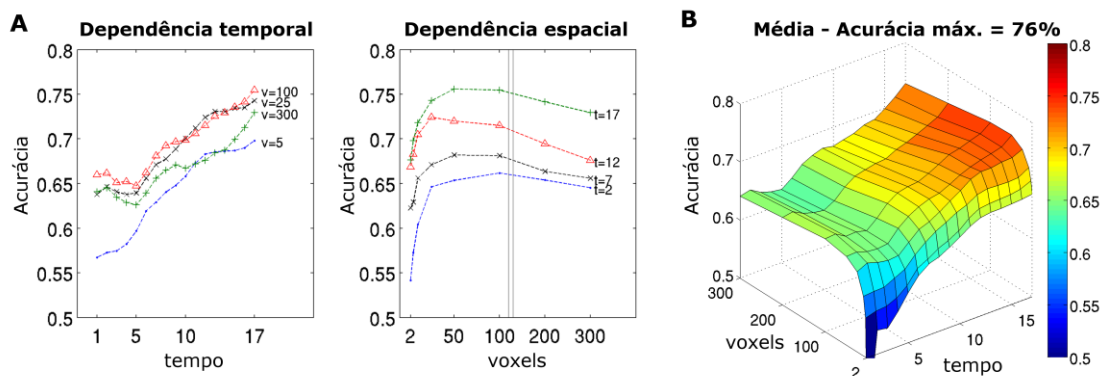


Figura 4-8: Acurácia média de seis sujeitos. (A) Dependência temporal (esquerdo) e espacial (direito) da acurácia. A acurácia do modelo aumenta de forma linear com o tempo enquanto satura em torno de 50 a 100 voxels depois de um aumento inicial forte. Nota-se que o tamanho do voxel é arbitrário e que tamanhos diferentes podem resultar em inclinações diferentes (B) A acurácia média em função do tempo e espaço mostra como o modelo melhora ao longo do tempo e como atinge melhores resultados com a inclusão de mais voxels.

Foi investigado quais regiões anatômicas contribuíram mais para a identificação. Os 300 *voxels* utilizados mais frequentemente para a identificação estão visualizados na Figura 4-9. Em cada iteração, os *voxels* foram selecionados de acordo com as correlações de predições do treino (validação cruzada interna). Os *voxels* mais utilizados, isto é, os *voxels* com a melhor posição se concentram no giro de Heschl e outras regiões auditivas secundárias como por exemplo o *planum temporale*, *planum polare* e giro temporal posterior e anterior. Outras regiões como o giro temporal médio, giro supramarginal e pólo temporal contribuíram menos para a identificação.

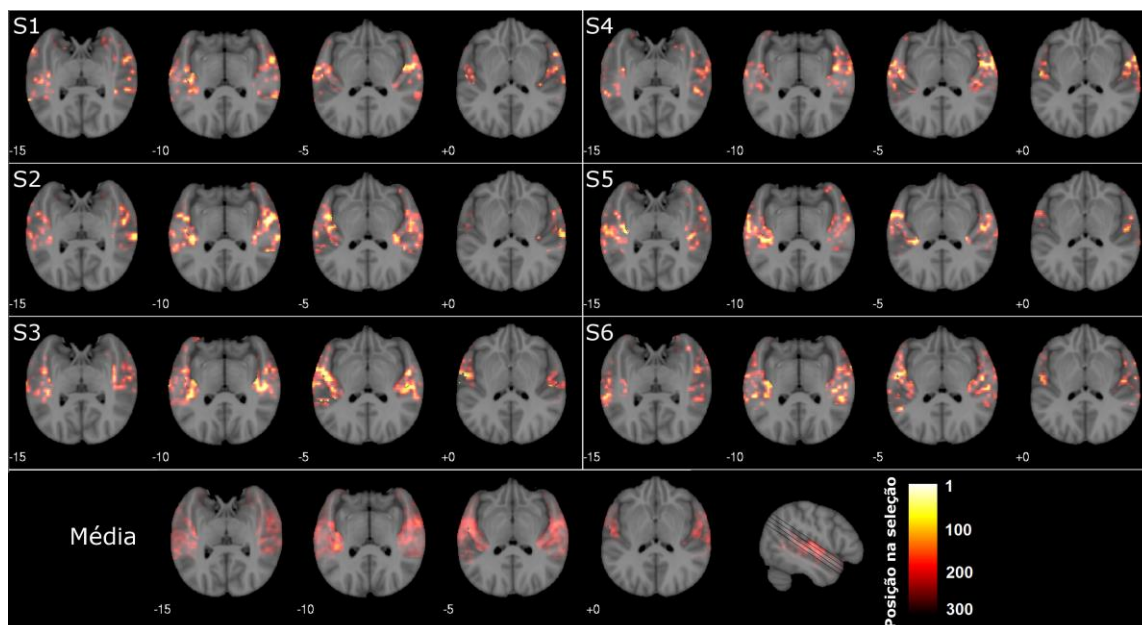


Figura 4-9: Voxels utilizados durante a identificação conforme a posição estabelecida durante o treino. São visualizados quatro cortes axiais com ângulo de  $34^\circ$  (PATTERSON; GRIFFITHS, 2002) para os seis sujeitos S1 a S6. A cor indica a posição obtida durante a estimativa do modelo com validação cruzada interna (primeiras posições: branco/amarelo, últimas posições: vermelho escuro). Embora haja diferenças individuais, os voxels nas primeiras posições se concentram no giro e Heschl, *planum temporale* e *planum polare*. O mapa indica que as características acústicas utilizadas envolvem processos auditivos primários e secundários representados no giro de Heschl e regiões anterior e posterior como *planum polare* e *planum temporale*, respectivamente.

#### ***Representações de características acústicas no córtex auditivo***

Para investigar possíveis subdivisões funcionais no córtex auditivo, foram calculadas as correlações entre cada característica acústica e a série temporal de cada um dos 300 melhores *voxels* separadamente para cada sujeito. Depois disso, os componentes principais dessa representação anatômica-funcional formalmente constituída pela matriz de 300 *voxels* por 21 correlações foram determinados com a análise de componentes principais (ACP) utilizando a função *princomp* do Matlab<sup>®</sup>. Essa análise deve resumir quais combinações de características acústicas explicam maior variância entre os *voxels*, isto é, quais subdivisões funcionais existem no córtex auditivo. Os resultados da ACP mostraram que a primeira componente principal (CP1) era extremamente semelhante entre os sujeitos (a correlação de *Pearson* entre todos os

pares de sujeitos ( $N = 6*5/2$ ):  $r=0.95\pm0.04$ ), e a segunda componente manteve ainda uma similaridade alta ( $r = 0.77\pm0.20$ ). Dessa forma, foi calculada uma ACP de grupo concatenando os *voxels* de todos os sujeitos. Os primeiros dois componentes explicaram 71% (48% + 23%) da variância. Como análise de controle, a ACP também foi calculada utilizando somente os melhores 100 *voxels*. Os componentes foram essencialmente os mesmos para esse número de *voxels* reduzido (similaridade dos primeiros dois componentes calculada pela correlação de Pearson comparando a versão de 100 e 300 *voxels* para cada sujeito ( $N=6$ ):  $r = 0.96\pm0.05$  e  $r = 0.86\pm0.17$  para CP1 e CP2, respectivamente). As subdivisões anátomo-funcionais dos componentes principais CP1 e CP2 e as combinações das características acústicas são visualizadas na Figura 4-10. Mesmos que uma certa variabilidade anatômica e funcional entre sujeitos é conhecida para essa região auditiva (MOEREL; DE MARTINO; FORMISANO, 2014), os resultados aqui mostraram uma clara preferência para frequências baixas no giro de Heschl representado pelos pesos positivos do CP1. Além dessa frequências baixa, o CP1 também representado um conceito chamado “Atividade” num estudo anterior (ALLURI et al., 2012). Pelo outro lado, o CP2 representa frequências mais altas e está localizado em regiões anteriores e posterior do giro de Heschl (HG), como por exemplo no *planum temporale* (PT) e *planum polare* (PP). Em resumo, a distribuição cortical mostra uma arquitetura de regiões principais, representada pelas frequências baixas localizadas no HG e frequências altas em PP e PT (MOEREL; DE MARTINO; FORMISANO, 2014).

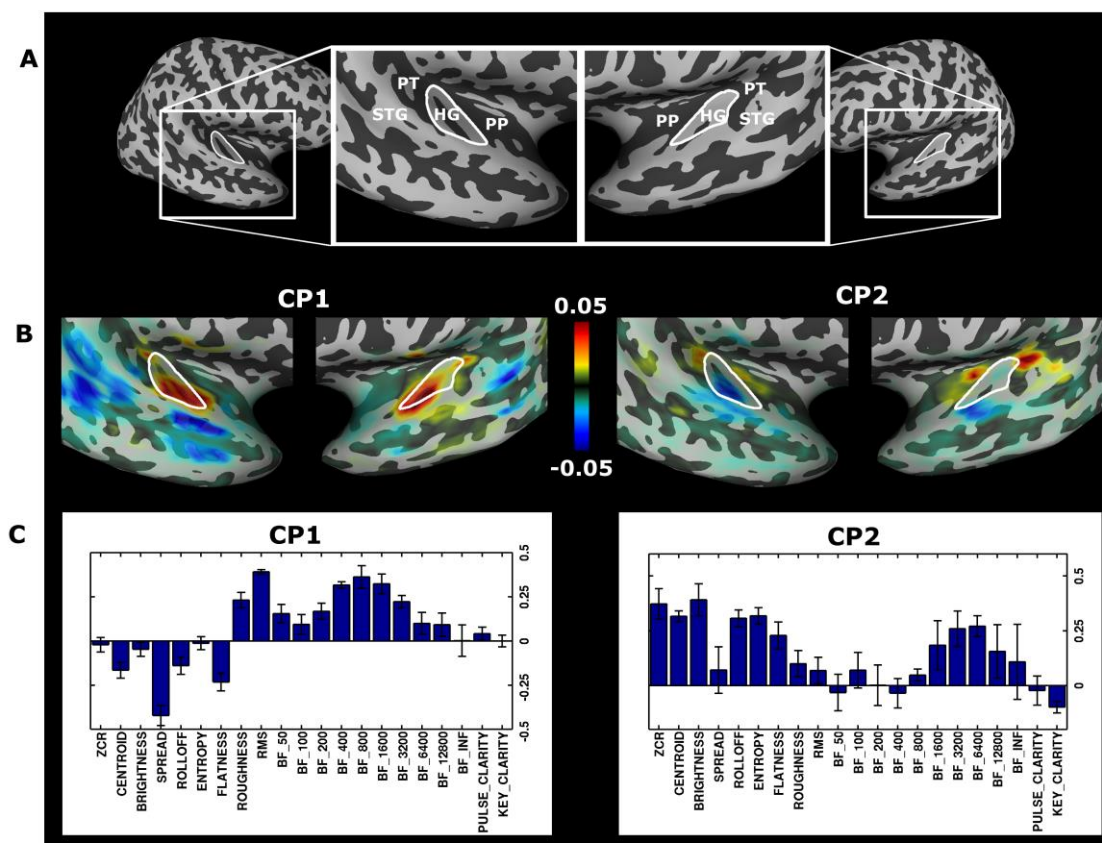


Figura 4-10: Distribuição cortical das características acústicas. (A) As seguintes regiões anatômicas são marcadas numa superfície inflada do córtex temporal de um sujeito exemplar: Giro de Heschl (HG), considerado córtex auditivo primário, planum polare (PP), planum temporale (PT) e giro temporal superior (STG) representando o córtex auditivo secundário. (B) A sobreposição do mapa colorido mostra os pesos dos dois primeiros componentes principais CP1 e CP2 no hemisfério direito e esquerdo (peso positivo: vermelho, peso negativo: azul; borda do HG traçado em branco). Os pesos foram multiplicados com as correlações do treino para destacar voxels responsivos a música. CP1 é localizado principalmente no HG, enquanto CP2 tem pesos positivo no PT e PP. (C) Carregamentos dos componentes CP1 e CP2 (média e desvio padrão dos sujeitos). CP1 tem carregamentos mais altos para as frequências baixas entre 200 Hz e 1600 Hz, representadas pelas características sub-band flux (BF com número indicando limite superior da faixa, ou seja, BF-400 representa frequências 200 Hz a 400 Hz). CP2 tem carregamentos positivos em frequências mais altas e brightness. Em resumo, os dois componentes principais destacam a preferência de frequências baixas do HG e de frequências altas das regiões próximas.

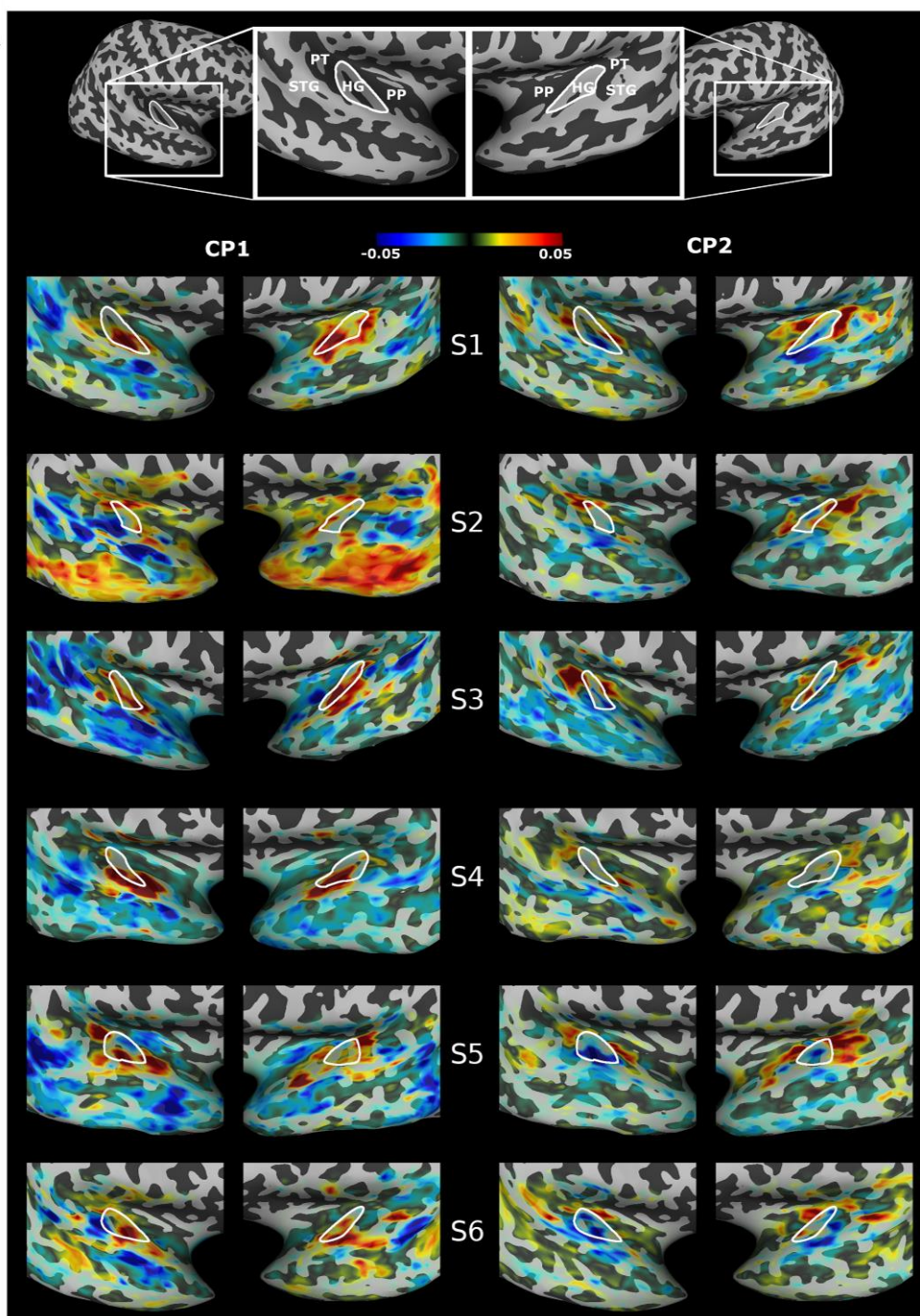


Figura 4-11: Distribuição cortical individual dos componentes. Além de diferenças individuais o componente CP1 tem representação mais próximo do giro de Heschl (HG) enquanto CP2 está anterior e posterior nas proximidades de planum polare (PP) e planum temporale (PT).



### ***Relação do desempenho do modelo e o número de variáveis***

Foi investigado se um modelo com menos variáveis possa atingir um desempenho equivalente na identificação das músicas. Por isso, o mesmo algoritmo de identificação foi utilizado, porém o número de características musicais no modelo foi sistematicamente aumentado de 1 a 21. Como não existe qualquer ordem intrínseca entre as características e para evitar de sub ou sob-estimar o efeito, aplicou-se esse procedimento duas vezes com ordem inversa do fator de inflação de variância, que mede o nível de colinearidade entre variáveis:

$$VIF = \text{diag}(\text{corr}(X))^{-1} \quad (11)$$

Os resultados mostram um aumento forte inicial da acurácia do modelo para as primeiras variáveis. Depois de aproximadamente 13 variáveis, a acurácia pouco aumenta (Figura 4-12). A relação entre o desempenho do modelo e o número de características acústicas foi levemente melhor explicada por uma função exponencial (no formato:  $a+b*exp(-cx)$ ) do que por uma função logarítmica (no formato  $a+b*ln(x)$ ). Os parâmetros foram estimados pela função de regressão não-linear *nlinfit* do Matlab® com os seguintes resultados:

Função exponencial:  $a = 0.77$ ;  $b = -0.30$ ;  $c = -0.18$

Função logarítmica:  $a = 0.51$ ;  $b = 0.09$

O coeficiente de determinação  $R^2$  foi calculado pela formula padrão, onde  $x_i$  é a acurácia obtida utilizando  $i$  variáveis,  $\hat{x}_i$  é o valor predito pela função estimada e  $\mu$  a média das acurácias:

$$R^2 = 1 - \frac{\sum (x_i - \hat{x}_i)^2}{\sum (x_i - \mu)^2} \quad (12)$$

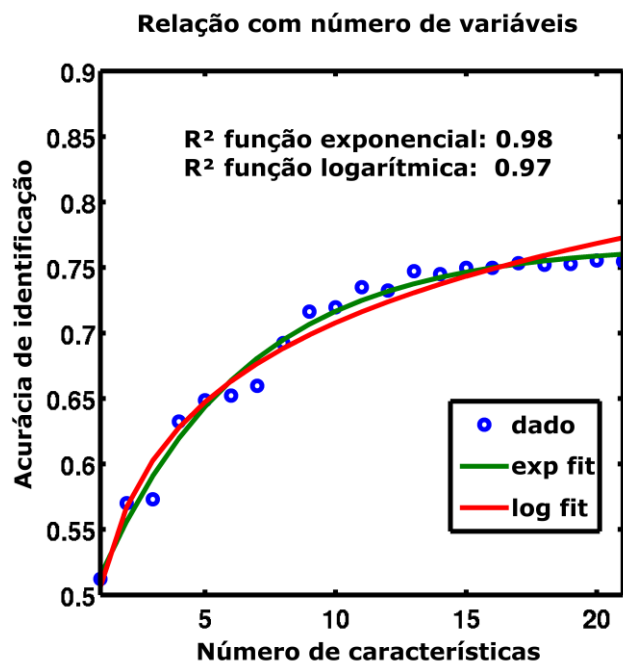


Figura 4-12: Relação da acurácia de identificação com o número de variáveis. A acurácia está próxima do nível de chance para um modelo com uma característica, aumenta substancialmente para modelos com aproximadamente 10-15 características e depois estabiliza. Pontos azuis representam a média sobre as duas ordens e sujeitos (utilizando 100 voxels e a duração completa da música). A curva verde representa a estimativa da função exponencial e a curva vermelha a função logarítmica.

#### **Robustez do modelo: Identificação de 10 músicas**

Foi investigada a robustez do modelo, estendendo o caso binário de identificação de 1 de 2 músicas para um número de músicas maior, isto é, deixando  $N = 3$  a  $N = 10$  músicas fora do treino. As mesmas músicas mantidas fora do treino foram, em seguida, decodificadas. O número de possíveis combinações de conjuntos de treino e teste aumenta drasticamente ( $N = 3$ : 9.880;  $N = 4$ : 91.390; ...). Para diminuir o tempo computacional, somente 10.000 combinações dos possíveis conjuntos foram testadas. Ainda, para o caso de  $N = 10$ , as 10 músicas do mesmo *medley*, isto é, todas as músicas de uma sequência funcional, foram reservadas para validação. Os resultados mostraram que a decodificação não é restrita para o caso binário. O desempenho do modelo manteve acurácias acima do nível de chance para acurácias percentuais ( $= 1/\#músicas$ ) e da mesma forma para acurácias posicionais, calculado por  $(N - \text{posição}(música\ predita))/(N - 1)$  (Figura 4-13).

No demais, a comparação entre a identificação de 10 músicas escolhidas aleatoriamente entre sequências e das 10 músicas da mesma sequência mostrou acurácias no mesmo nível ( $73,8\% \pm 5,3\%$  em comparação com  $74,4\% \pm 6,0\%$ ; teste  $t$  pareado  $t(5) = -0,86$ ,  $p = 0,42$ ; conferir Figura 4-14). Esse resultado demonstra que o algoritmo não é só estável para a identificação de várias músicas como também para um caso extremo no qual todas as músicas de teste foram escutadas numa sequência funcional separada, isto é, num dado de teste temporalmente independente. Isto indica que não existe qualquer viés nas acurácias relatadas que, possivelmente, pudesse ser introduzido por correlações temporais entre músicas de uma mesma sequência (VAROQUAUX et al., 2017).

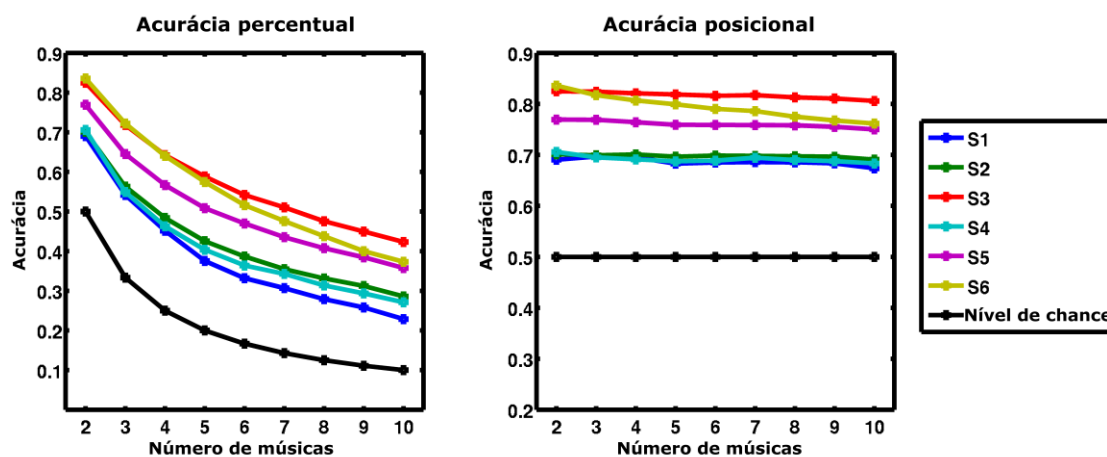
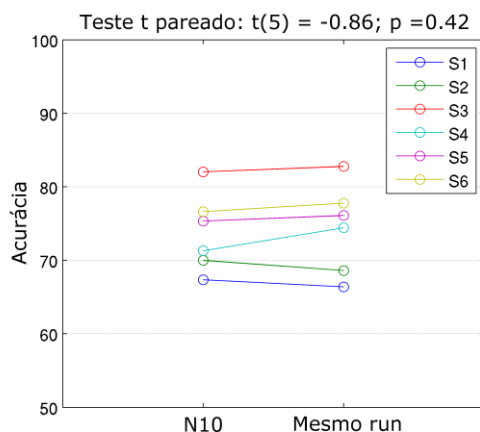


Figura 4-13: Robustez do modelo na identificação de 2 a 10 músicas. Acurácias se mantinham acima do nível de chance para todos os sujeitos S1 a S6 para a métrica original, a acurácia percentual (lado esquerdo). Da mesma forma, a acurácia posicional, calculado por  $(N - \text{posição(música predita)}) / (N - 1)$  se mantinha constantemente no mesmo nível. Esse resultado mostra que a abordagem de decodificação não está restrita para o caso binário.



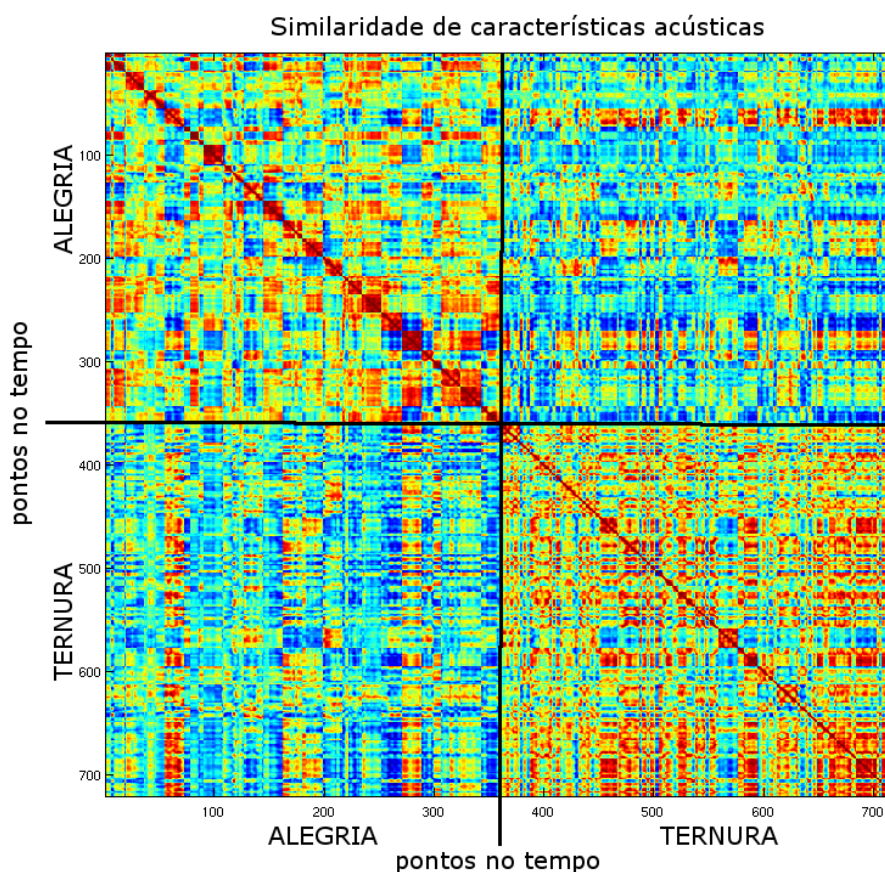
*Figura 4-14: Comparação de identificação de 10 músicas escolhidas aleatoriamente entre as sequências e da mesma sequência funcional. Essa análise de controle compara a acurácia entre escolhas aleatórias de 10 de 40 músicas e a escolha de 10 músicas do mesmo medley, isto é, da mesma sequência funcional. As acurácias de identificação são equivalentes e não mostram qualquer diferença sistemática entre essas duas formas de validação (teste t pareado  $t(5) = -0.86$ ,  $p=0.42$ ). Esse resultado indica que as acurácias não sofrem qualquer viés que poderia ser possivelmente introduzido por correlações temporais entre músicas de treino e teste dentro.*

### ***Entropia de Shannon***

Numa última análise, a entropia de Shannon foi utilizada para explicar parte da variação das acurácias. Essa análise se baseou na hipótese de que músicas com um grau de informação maior, possivelmente, podem ser identificadas mais precisamente. Para isto, a entropia de Shannon foi calculada a partir das características acústicas para cada música  $X$  da seguinte forma. Primeiro, a similaridade entre as características acústicas foi calculada pela correlação de Pearson entre todos os possíveis pares de pontos no tempo. Ou seja, foi calculada a matriz de correlação sobre os pontos do tempo da música (18 – já que 5 volumes da transição foram descartados) e número de músicas (40), isto é, uma matriz de  $720 \times 720$  (Figura 4-15). A faixa de todos os valores de similaridade foi subdividida em 10 intervalos equidistantes. Esses intervalos foram utilizados para calcular a entropia de Shannon (SHANNON, 1948) para cada música  $X$ , utilizando todos os pares de pontos no tempo fazendo parte da música:

$$H(X) = - \sum_{i=1}^{10} p_i \log_2 p_i$$

onde  $p_i$  é a frequência do valor de similaridade entre pontos no tempo da música  $X$  no intervalo  $i$ . Em resumo, essa métrica reflete a variabilidade entre os valores de similaridade entre características acústicas entre todos os momentos da mesma música.



*Figura 4-15: Similaridade de características acústicas entre todos os pontos no tempo de todas as músicas. As músicas foram concatenadas antes de calcular a similaridade. O quadrante do esquerdo superior representa a similaridade entre as 20 músicas da categoria emocional ALEGRIA, o quadrante da direita inferior da TERNURA. Destacam se esses dois quadrantes visualmente por ter maior semelhança acústica dentro da categoria do que entre as categorias (quadrantes da direita superior e esquerda inferior). De forma similar, a maioria das músicas se destaca visualmente (quadrados vermelhos ao longo da diagonal da matriz) por possuir similaridade maior dentro de cada uma das músicas do que entre outras músicas.*

Para investigar o efeito dessa métrica na acurácia, os dados de validação foram divididos em conjuntos de músicas com entropia alta e baixa. Os resultados mostraram que essa divisão, de fato, tem um efeito forte no desempenho do modelo (Figura 4-16, lado esquerdo). A acurácia calculada sobre os 45 pares das 10 músicas com entropia alta atingiu 95,5% para três sujeitos, 93,3%, 90,0% e 76,7% para S3, S2 e S1 ( $91\% \pm 7\%$ ; média $\pm$ DP entre sujeitos) em comparação a  $71\% \pm 12\%$  (média $\pm$ DP entre sujeitos) para músicas com entropia baixa. Este mesmo efeito foi confirmado quando a divisão entre músicas de entropia alta e baixa foi realizada em cada categoria emocional separadamente (Figura 4-16, lado direito).

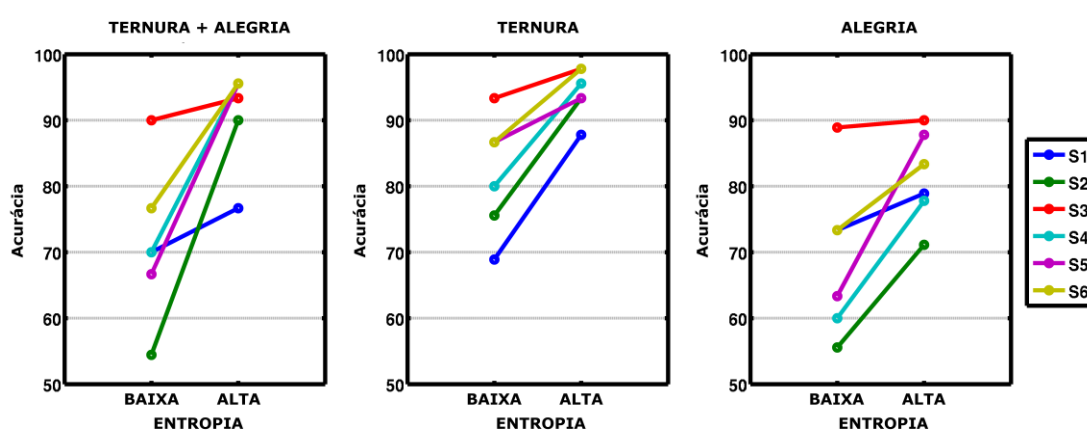


Figura 4-16: Entropia como fator importante na acurácia: os dados de validação foram divididos em conjuntos de 10 músicas com baixa e alta entropia. O efeito da entropia na acurácia se confirmou quando a divisão utilizou todas as músicas (lado esquerdo), somente músicas de ternura (meio) ou alegria (lado direito). Uma ANOVA com os dois fatores de conjunto de validação (TERNURA+ALEGRIA, TERNURA, ALEGRIA) e entropia (BAIXA, ALTA) confirmou os efeitos principais para conjunto de validação ( $F(2;10)=15,3$ ,  $p<0,0009$ ) e entropia ( $F(1;5) = 23,81$ ,  $p=0,004$ )

## 5 Discussão

Na análise principal, foi investigada a resposta cerebral a 40 músicas de vários estilos com um modelo de codificação e decodificação. Esta abordagem da combinação da codificação com a decodificação estende abordagens já existentes (ALLURI et al., 2013; TOIVAINEN et al., 2014) e corrobora a importância de utilizar múltiplas variáveis para a identificação. Nesse modelo, as características acústicas descrevem a música ao longo do tempo e permite localizar as áreas cerebrais do córtex auditivo que são críticas para a identificação de músicas. Essa abordagem foi capaz de identificar músicas com aumento de acurácia ao longo do tempo e espaço, demonstrando a dependência espaço-temporal da acurácia para a identificação de músicas. A informação distribuída no córtex auditivo junto com a entropia de músicas atingiu uma acurácia máxima de 95%. Vale destacar que esses resultados foram alcançados para um conjunto heterogêneo de músicas (música clássica, rock, pop, jazz e música popular com e sem letras), representando distintas categorias emocionais como ternura e alegria. Os resultados estão publicamente disponíveis em [institutor.github.io/mirviewer/SI](https://institutor.github.io/mirviewer/SI) para promover um entendimento maior da organização espaço-temporal de representações acústicas complexas. Espera-se que, futuramente, essa contribuição ajude no desenvolvimento de interfaces de cérebro-máquina na área de neuromodulação em diversos cenários, como, por exemplo, no tratamento de pacientes sofrendo de alucinações auditivas (DIERKS et al., 1999). Uma outra possível aplicação consiste na reconstrução de música imaginada utilizando um banco de dados grande de músicas formalmente como probabilidade *a priori*, em analogia a estudos de reconstrução de conteúdo visual de imagens (NASELARIS et al., 2009).

### **Ponto de inflexão da acurácia como limite de significância prática**

Em comparação a estudos anteriores que se concentraram ou na codificação ou na decodificação separadamente, esse estudo combinou uma segunda etapa de decodificação após a primeira etapa de codificação. Essa combinação permite não somente identificar músicas novas (isto é, músicas não utilizadas durante a etapa de codificação), como também investigar explicitamente como as dimensões espaço e tempo se relacionam com a acurácia do modelo. Foi confirmada a relação já esperada

entre tempo e acurácia de identificação. Para a dimensão espacial, ao se adicionar mais *voxels* durante a etapa de decodificação, foi possível determinar até que ponto o modelo se beneficiou dessa abordagem multivariada. Interessantemente, depois de um aumento inicial forte da acurácia até uns 50-100 *voxels* (tamanho do *voxel*  $3,75 \times 3,75 \times 4,75 = 66 \text{ mm}^3$ ), o desempenho do modelo saturou e até piorou ao se incluírem mais *voxels*. Uma explicação plausível é que as características acústicas não podem explicar o sinal desses *voxels* e, conseqüentemente, não existe uma predição confiável que possa contribuir para a identificação. Ao contrário, essa predição “nula” introduz mais ruído no modelo e, por isso, diminui a acurácia. Este ponto de inflexão do desempenho, isto é, o ponto onde a acurácia termina de aumentar e começa a declinar, pode ser aproveitado para determinação de significância prática em oposição a testes estatísticos clássicos. Em vez de escolher um valor-*p* arbitrário como limiar estatístico (BAKER, 2016), aqui, se sugere considerar somente os *voxels* significantes que contribuem na prática para o desempenho do modelo. Em resumo, é a combinação das duas etapas no modelo junto com a investigação sistemática que promove a determinação da significância prática que vai além de testes estatísticos clássicos.

### **Robustez e a identificação de músicas de uma sessão independente**

As acurácias absolutas encontradas neste estudo são comparáveis ou até maiores do que em estudos anteriores nos domínios auditivos (DE MARTINO et al., 2008; SANTORO et al., 2014) e de música (TOIVIAINEN et al., 2014). Uma publicação recente apresentou acurácias robustas acima do nível de chance (CASEY, 2017). Porém, a análise de decodificação consistiu em calcular a similaridade de respostas cerebrais entre apresentações repetidas do mesmo estímulo sem construir um modelo de codificação. Além dessa diferença metodológica principal, a comparação de acurácias absolutas não é simples já que o dado do estudo citado acima foi adquirido numa RM de alto-campo de 7T, e o procedimento da validação cruzada, diferentemente do presente estudo, foi realizada separando dado de treino e teste do mesmo sujeito. Numa publicação recente, Varoquaux et al. (2016) advertiu que esse tipo de validação cruzada poderia gerar acurácias enviesadas, especialmente quando observações individuais (isto é, dados de volumes da mesma sequência de RMf) fazem parte do treino e teste. Por outro lado, a abordagem do presente estudo mantém um bloco de dado (a música inteira de 46 s) em vez de observações individuais (volumes) e comprovou-se a robustez para a



identificação de até 10 músicas. É importante destacar que a acurácia para identificação de todas músicas de uma mesma sequência não utilizada para o treino se manteve no mesmo nível (Figura 4-14: Comparação de identificação). Esta robustez indica que, nesse estudo, as acurácias não sofrem da validação cruzada com dado do mesmo sujeito.

### **Número de variáveis e extensão de características acústicas**

A investigação da relação do desempenho do modelo com o número de variáveis confirmou a importância de ter uma descrição completa do espaço dos estímulos, aqui em forma de múltiplas características acústicas. Mesmo que a acurácia tenha aumentado monotonicamente, a relação foi melhor explicada por uma função exponencial do que por uma função logarítmica. Isso poderia indicar que o modelo, mesmo com mais características acústicas, não melhoraria o desempenho. Porém, esse resultado vale para esse conjunto específico e homogêneo utilizado nesse estudo. Estudos futuros devem explorar outras categorias de características acústicas, como, por exemplo, descritores de harmonia, melodia e similaridade. Outros possíveis descritores podem ser consultados na ampla literatura no campo de recuperação de informações de música (COUTINHO; CANGELOSI, 2011; GINGRAS; MARIN; FITCH, 2014; LAURIER et al., 2009). Porém, as escolhas não são simples e incluem decisões sobre tamanho de janelas e estatísticas sobre intervalos menores ou maiores para a integração temporal.

### **Generalização entre gêneros de música**

A análise apresentada aqui tem alguns aspectos em comum com a validação cruzada de Toiviainen et al. (2014), que classifica segmentos de música com duração variada. Porém, a música representava apenas um estilo (lado B do álbum *Abbey Road*, *The Beatles*) e a variável dependente era um componente acústico. Aqui, as músicas representaram diversos estilos como música clássica, rock, pop, jazz e música popular. Além disso, as músicas induziram emoções como ternura e alegria que diferiam bastante nas características acústicas (Figura 4-15: Similaridade de características acústicas). Mesmo que tenha sido demonstrado que diferentes estilos de músicas induzem padrões cerebrais distintos, a abordagem elaborada aqui era capaz de generalizar entre os gêneros e estilos de música. Isso indica que o modelo representa,

implicitamente, diferenças nos gêneros, o que é um ponto forte para a predição de músicas novas de outros gêneros.

### **Entropia de Shannon**

A investigação da acurácia e da entropia das músicas foi motivada pela observação que as músicas tinham graus de variações diferentes. Intuitivamente, quanto maior variação a música possui, mais chance de ter uma boa predição da música. Matematicamente, a predição é calculada pela correlação entre resposta cerebral predita e adquirida. Por isso, considerando a dimensão do tempo, a predição necessariamente precisa de variação temporal. No caso extremo de nenhuma variação temporal, a predição seria constante e somente a dimensão espacial poderia contribuir para identificação. Em resumo, a informação temporal classificada pela entropia de Shannon é um fator importante para uma identificação precisa de músicas.

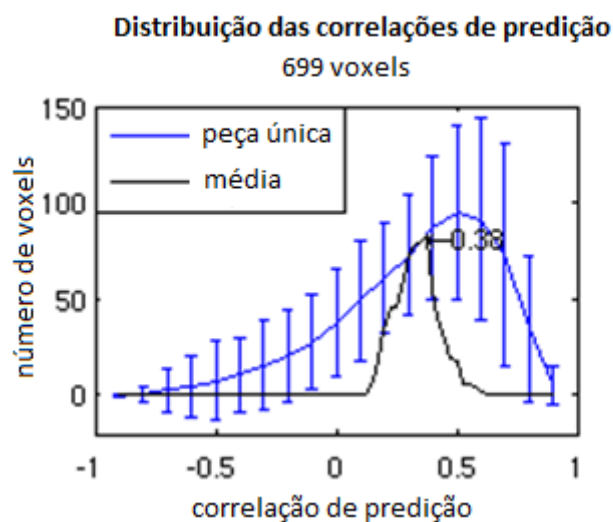
### **Representações funcionais de características acústicas**

O modelo de codificação permitiu determinar a representação neural das características acústicas. O giro de Heschl, regiões anteriores e posteriores, como *planum polare* e *planum temporale*, e partes do giro temporal superior foram as áreas melhor preditas. Esse resultado é consistente com o estudo de Norman-Haignere et al. (2015), que demonstrou áreas anteriores e posteriores do giro temporal superior como as mais responsivas para música em comparação a outros sons de ambiente, de natureza e de fala. Além disso, as regiões encontradas nesse trabalho foram relatadas e investigadas em relação a música em diversos estudos (GRIFFITHS et al., 1998; GRIFFITHS; WARREN, 2002; LEAVER; RAUSCHECKER, 2010; PATTERSON; GRIFFITHS, 2002; WARREN et al., 2003). Ainda, a análise de componentes principais revelou uma componente com preferência de frequências baixas e outra com preferência de frequências altas. A representação desses componentes é consistente com numerosos estudos de tonotopia (FORMISANO et al., 2003; HERDENER et al., 2013; MOEREL; DE MARTINO; FORMISANO, 2012, 2014; SAENZ; LANGERS, 2014; SCHÖNWIESNER et al., 2014; SCHÖNWIESNER; VON CRAMON; RÜBSAMEN, 2002; THOMAS et al., 2015). Partindo desses resultados consistentes, estudos futuros

deveriam investigar outras combinações de características acústicas que poderiam revelar novas dimensões e divisões funcionais no córtex auditivo.

### Limitação técnica e experimental

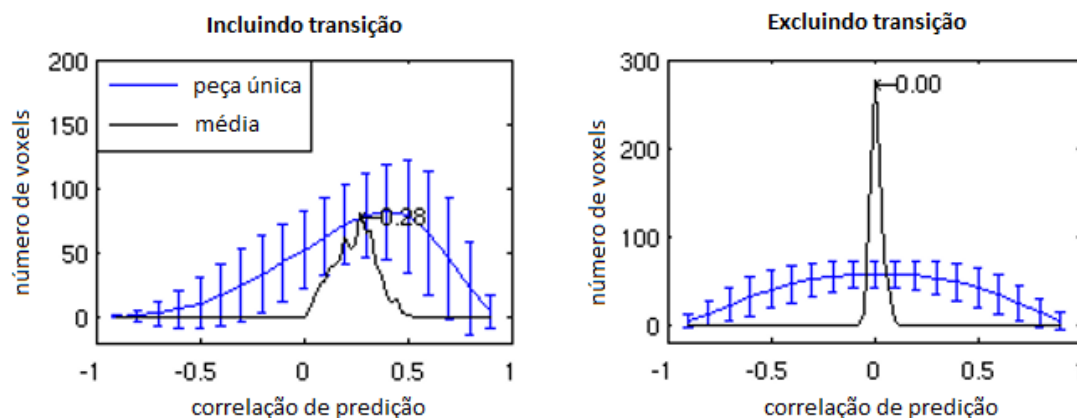
Uma limitação da acurácia de forma geral foi observada quando se visualizou a distribuição das correlações de predição entre os *voxels*. Existe uma grande variabilidade na medida da correlação de predição quando essa é calculada sobre somente uma música, isto é, sobre 23 pontos. Esta variabilidade se reduz drasticamente quando se calcula a média sobre todas as músicas. Esta observação é exemplificada com um sujeito na Figura 5-1.



*Figura 5-1: A distribuição das correlações de predição mostra grande variabilidade para as correlações de somente uma música (em azul; barra mostra o desvio padrão sobre 40 músicas). Obtém-se uma grande quantidade de voxels com correlações negativas como também correlações extremamente positivas. Após de calcular a média sobre todas as predições de todas as 40 músicas, o intervalo das correlações diminui e se concentra em valores com correlações positivas moderadas.*

Uma outra limitação foi encontrada quando a análise incluía os períodos de transição entre as músicas, realizado na construção do *medley* por um *fade-out* e *fade-in* de 1 s. Esperar-se-ia uma distribuição simétrica com média zero, quando se utilizam as características de uma música não-correspondente, ou seja, na predição da música **A** com características de uma outra música **B**. Porém, quando se inclui a transição entre as

músicas, essa parte comum entre todas as músicas pode resultar em predições positivas mesmo quando se usam características não-correspondentes, como mostra a *Figura 5-2*. Devido a essa observação, a análise de identificação descartou os primeiros 6 volumes.



*Figura 5-2: Correlações de predição para características não-correspondentes: esperam-se predições com média zero para este caso. Porém, predições positivas são observadas quando inclui a transição entre músicas na análise (lado esquerdo). Excluindo a transição (descartando os primeiros 6 volumes), as predições estão centralizadas por volta do zero, como esperado.*

O resultado principal da identificação de músicas corrobora a possibilidade de prever e decodificar estímulos a partir da atividade cerebral. As acurácias obtidas são comparáveis com outros estudos como, por exemplo, o de SANTORO et al. (2014), que identificou sons de diferentes tipos, e o de TOIVIAINEN et al. (2014), que identificou segmentos de uma música. Embora existam várias diferenças entre os estudos (duração e tipo de estímulos), inclusive no presente estudo, as acurácias não se comparam com os resultados obtidos na área visual. SANTORO et al. (2014) leva em consideração que o contraste funcional do sinal BOLD para o domínio auditivo é menor, uma vez que o equipamento da ressonância magnética apresenta um ruído acústico. Outra possível causa é o modelo em si e também o número de variáveis e pontos. Observou-se, em análises prévias, neste projeto a importância de se ter um número mínimo de pontos disponíveis para estimar a resposta cerebral de forma robusta. Não só a estimativa é sujeita ao ruído. Mesmo com um modelo correto e uma estimativa adequada que reflete a resposta verdadeira do sinal, a predição pode sofrer com o ruído intrínseco de qualquer medição de um sinal biológico. Esse fato é implicitamente representado nos

resultados desse estudo, que mostra acurácias altas somente a partir de certo número de volumes considerados para a identificação. O mesmo ponto, muitas vezes, também é considerado nos desenhos de estudos. Como levantado na seção 2.4, vários estudos repetiram o mesmo estímulo na sessão de validação, enquanto existe a preferência de ter um dado para o treinamento o mais variado possível com apenas uma apresentação.

Ainda no contexto da modelagem de um sinal ruidoso com pouco contraste funcional, o presente trabalho também mostrou a importância de cada passo no pré-processamento para preservar e aproveitar ao máximo o sinal da atividade cerebral. Além da subtração de efeitos fisiológicos, a remoção do sinal global se mostrou especialmente vantajosa para a modelagem. Possivelmente, outras melhorias poderiam ser consideradas no futuro, como, por exemplo, a estimativa do formato da resposta hemodinâmica por característica musical. E, por fim, as escolhas ideais para tamanhos de filtros espaciais e temporais ainda puderam ser investigadas mais a fundo. Na prática, porém, a avaliação de cada passo e parâmetro apresenta uma explosão combinatória que nem sempre pode ser respondida de forma exaustiva.

Na análise principal, os dados foram agregados ignorando os fatores experimentais emoção e tarefa. O sucesso da identificação mesmo com a agregação pode ser um indício de que o efeito desses fatores na atividade do córtex auditivo é pequeno. Porém, também é possível que o conjunto das características acústicas não tenha sido sensível o suficiente para constatar uma diferença no mapeamento. Outro indício que aponta na mesma direção é a falta de consistência na predição de regiões não-auditivas: análises prévias mostraram que as melhores predições se concentraram no córtex temporal. Outras regiões mais frontais obtiveram predições inferiores e não consistentes entre indivíduos e condições experimentais. Correlações de BOLD a BOLD entre repetições da mesma condição experimental também mostraram pouca consistência, tanto dentro do mesmo indivíduo como entre indivíduos. Essa falta de consistência na resposta cerebral entre repetições do mesmo estímulo inviabiliza a modelagem com características acústicas, a qual assume um sinal idêntico entre repetições. Vários motivos podem contribuir para um sinal pouco consistente. Já foi demonstrado que regiões frontais possuem características mais individuais do que regiões sensoriais (FINN et al., 2015) o que sugere uma variação funcional maior entre indivíduos. Outras variáveis importantes individuais são processos cognitivos internos como, por exemplo, introspecção ou *mind wandering* (uma mente vagando) que podem

ocorrer ao ouvir música (MASON et al., 2007; TROST et al., 2011). Esses processos desacoplados ao estímulo musical em si podem influenciar a atividade cerebral, são difíceis de controlar e registrar em tempo real, e, por isso, apresentam uma limitação na modelagem que requer um acoplamento direto de características acústicas com a atividade cerebral. Eventualmente, a própria tarefa experimental de sentir emoção ou analisar a música tecnicamente pode recrutar redes neurais que tenham uma atividade própria e igualmente desacoplado ao estímulo musical. Esses fatores adicionais apresentam uma limitação inerente a qualquer tipo de modelagem. Possivelmente, estudos futuros podem considerar esses fatores e tentar desenvolver ferramentas para separar melhor os diversos processos cognitivos que ocorrem simultaneamente.

## 6 Conclusão

Este trabalho investigou as respostas cerebrais à música com um protocolo experimental naturalístico. Respostas consistentes a características acústicas das músicas foram encontradas no córtex auditivo, em especial no giro de Heschl, *planum temporale* e *planum polare*. Baseado nessas regiões, foi factível prever e identificar músicas com uma acurácia de até 95% para aquelas com maior entropia. Vale destacar que esses resultados foram obtidos com validação cruzada, uma técnica robusta oriunda da área de aprendizagem de máquina. Além disso, a modelagem de codificação com as características acústicas revelou detalhadamente a organização cortical do processamento auditivo. Em resumo, essa análise avaliou vários parâmetros internos e externos do modelo com o intuito de otimizar e aperfeiçoar o algoritmo de decodificação de músicas.

Os resultados promissores obtidos nesse estudo para a identificação de novas músicas a partir de características acústicas abrem possibilidades para futuros desenvolvimentos e aplicações. Por exemplo, o processo de identificação de músicas poderia ser reformulado para um processo de reconstrução de músicas. Para isto, o modelo precisaria incluir um *prior* em forma de um banco de dados com um grande número de músicas e criar uma reconstrução utilizando as músicas com maior similaridade a partir da atividade cerebral observada. Isto, por sua vez, poderia ser incorporado numa aplicação clínica de neurofeedback, por exemplo, no tratamento de alucinações auditivas que tem atividade neuronal comprovada. De forma similar, a investigação do padrão de conectividade na tarefa de sentir a emoção poderia ser futuramente utilizada como referência para aplicações de neurofeedback, eventualmente apoiando o tratamento de outros transtornos como a depressão.

## 7 Referências bibliográficas

ABDEL-GHAFFAR, S. et al. “Using voxel-wise encoding models to study occipito-temporal representations of the animacy, semantic and affective content of natural images.” **Journal of vision**, v. 15, n. 12, p. 508, 1 set. 2015.

ALLURI, V. et al. “Large-scale brain networks emerge from dynamic processing of musical timbre, key and rhythm.” **NeuroImage**, v. 59, n. 4, p. 3677–89, 15 fev. 2012.

ALLURI, V. et al. “From Vivaldi to Beatles and back: predicting lateralized brain responses to music.” **NeuroImage**, v. 83, p. 627–36, dez. 2013.

ALLURI, V.; TOIVAINEN, P. “Exploring Perceptual and Acoustical Correlates of Polyphonic Timbre”. **Music Perception**, v. 27, n. Escm, p. 223–242, 2010.

AUST, S. et al. “Experience in Remitted Depression”. v. 3, n. 2, p. 1–10, 2013.

BAKER, M. “Statisticians issue warning over misuse of P values”. **Nature**, v. 531, n. 7593, p. 151–151, 7 mar. 2016.

BAUMANN, S.; PETKOV, C. I.; GRIFFITHS, T. D. “A unified framework for the organization of the primate auditory cortex”. **Frontiers in Systems Neuroscience**, v. 7, n. April, p. 1–8, 2013.

BIRN, R. M. et al. “Separating respiratory-variation-related fluctuations from neuronal-activity-related fluctuations in fMRI”. **NeuroImage**, v. 31, p. 1536–1548, 2006.

BIRN, R. M. et al. “The respiration response function: The temporal dynamics of fMRI signal fluctuations related to changes in respiration”. **NeuroImage**, v. 40, n. 2, p. 644–654, 2008.

BLOOD, A J.; ZATORRE, R. J. “Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion.” **Proceedings of the National Academy of Sciences of the United States of America**, v. 98, n. 20, p. 11818–23, 25 set. 2001.

CASEY, M. A. “Music of the 7Ts: Predicting and Decoding Multivoxel fMRI Responses with Acoustic, Schematic, and Categorical Music Features”. **Frontiers in Psychology**, v. 8, n. July, p. 1–11, 2017.

CHAI, X. J. et al. “Anticorrelations in resting state networks without global signal regression.” **NeuroImage**, v. 59, n. 2, p. 1420–8, 16 jan. 2012.

CHAPIN, H. et al. “Dynamic emotional and neural responses to music depend on performance expression and listener experience”. **PLoS ONE**, v. 5, n. 12, p. 5, 2010.

CHEN, G. et al. “A method to determine the necessity for global signal regression in resting-state fMRI studies.” **Magnetic resonance in medicine**, v. 68, n. 6, p. 1828–35, dez. 2012.

CHIKAZOE, J. et al. “Population coding of affect across stimuli, modalities and individuals.” **Nature neuroscience**, v. 17, n. 8, p. 1114–1122, 2014.

COUTINHO, E.; CANGELOSI, A. “Musical emotions: Predicting second-by-second subjective feelings of emotion from low-level psychoacoustic features and



physiological measurements.” **Emotion (Washington, D.C.)**, v. 11, n. 4, p. 921–937, 2011.

ÇUKUR, T. et al. “Attention during natural vision warps semantic representation across the human brain.” **Nature neuroscience**, v. 16, n. 6, p. 763–70, jun. 2013.

DAMADIAN, R. “Tumor Detection by Nuclear Magnetic Resonance”. **Science**, v. 171, n. 3976, 1971.

DAMADIAN, R.; GOLDSMITH, M.; MINKOFF, L. “NMR in cancer: XVI. FONAR image of the live human body.” **Physiological chemistry and physics**, v. 9, n. 1, p. 97–100, 108, 1977.

DE MARTINO, F. et al. “Combining multivariate voxel selection and support vector machines for mapping and classification of fMRI spatial patterns.” **NeuroImage**, v. 43, n. 1, p. 44–58, 15 out. 2008.

DIERKS, T. et al. “Activation of Heschl’s Gyrus during Auditory Hallucinations”. **Neuron**, v. 22, p. 615–621, 1999.

FINN, E. S. et al. “Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity”. **Nature Neuroscience**, n. October, p. 1–11, 2015.

FORMISANO, E. et al. “Mirror-Symmetric Tonotopic Maps in Human Primary Auditory Cortex”. **Neuron**, v. 40, n. 4, p. 859–869, 2003.

FOX, M. D. et al. “The global signal and observed anticorrelated resting state brain networks.” **Journal of neurophysiology**, v. 101, p. 3270–3283, 2009.

GABRIELSSON, A.; WIK, S. “Strong experiences related to music: A descriptive system.” **Musicae Scientiae**, 2003.

GERLACH, W.; STERN, O. “Das magnetische Moment des Silberatoms”. **Zeitschrift für Physik**, v. 9, n. 1, p. 353–355, dez. 1922.

GINGRAS, B.; MARIN, M. M.; FITCH, W. T. “Beyond intensity: Spectral features effectively predict music-induced subjective arousal.” **Quarterly journal of experimental psychology (2006)**, v. 67, n. July, p. 1428–46, 2014.

GLASSER, M. F. et al. “The Human Connectome Project’s neuroimaging approach.” **Nature neuroscience**, v. 19, n. 9, p. 1175–87, 2016.

GLOVER, G. H.; LI, T. Q.; RESS, D. “Image-based method for retrospective correction of physiological motion effects in fMRI: RETROICOR.” **Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine**, v. 44, n. 1, p. 162–7, jul. 2000.

GREEN, S. B. “How Many Subjects Does It Take To Do A Regression Analysis”. **Multivariate Behavioral Research**, v. 26, n. 3, p. 499–510, jul. 1991.

GRIFFITHS, T. D. et al. “Analysis of temporal structure in sound by the human brain.” **Nature neuroscience**, v. 1, n. 5, p. 422–7, 1998.

GRIFFITHS, T. D.; WARREN, J. D. “The planum temporale as a computational hub”. **Trends in Neurosciences**, v. 25, n. 7, p. 348–353, 2002.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. “The Elements of Statistical Learning”. **Elements**, v. 1, p. 337–387, 2009.

HERDENER, M. et al. “Spatial representations of temporal and spectral sound

cues in human auditory cortex”. **Cortex**, v. 49, n. 10, p. 2822–2833, 2013.

HUETTEL, S. A.; SONG, A. W.; MCCARTHY, G. **Functional Magnetic Resonance Imaging**. 2nd. ed. Massachusetts: Sinauer Associates, 2004.

HUMPHRIES, C.; LIEBENTHAL, E.; BINDER, J. R. “Tonotopic organization of human auditory cortex.” **NeuroImage**, v. 50, n. 3, p. 1202–11, 15 abr. 2010.

HUTH, A. G. et al. “A Continuous Semantic Space Describes the Representation of Thousands of Object and Action Categories across the Human Brain”. **Neuron**, v. 76, p. 1210–1224, 2012.

JAMES, W. “II. - What is an emotion?” **Mind**, v. 9, p. 188–205, 1884.

JENKINSON, M. et al. “Improved optimization for the robust and accurate linear registration and motion correction of brain images”. **NeuroImage**, v. 17, n. 2, p. 825–841, 2002.

JUSLIN, P. N. “From everyday emotions to aesthetic emotions: Towards a unified theory of musical emotions”. **Physics of Life Reviews**, v. 10, p. 235–266, 2013.

KANDEL, E. R.; SCHWARTZ, J. H.; JESSELL, T. M. **Principles of neural science**. 4 edition ed. New York: McGraw-Hill Medical, 2000.

KASPER, L. et al. “Cardiac artefact correction for human brainstem fMRI at 7 Tesla”. **Proc Org Hum Brain Mapp**, v. 15, p. 395, 2009.

KASPER, L. “PhysIO Toolbox Manual”. p. 1–22, 2013.

KAY, K. N. et al. “Identifying natural images from human brain activity.” **Nature**, v. 452, n. March, p. 352–355, 2008.

KOELSCH, S. “A neuroscientific perspective on music therapy.” **Annals of the New York Academy of Sciences**, v. 1169, p. 374–84, jul. 2009.

KOELSCH, S. “Towards a neural basis of music-evoked emotions.” **Trends in cognitive sciences**, v. 14, n. 3, p. 131–7, mar. 2010.

KOELSCH, S. “Toward a neural basis of music perception - a review and updated model.” **Frontiers in psychology**, v. 2, 2011.

KOELSCH, S. et al. “The roles of superficial amygdala and auditory cortex in music-evoked fear and joy”. **NeuroImage**, v. 81, p. 49–60, 2013.

KOELSCH, S. “Brain correlates of music-evoked emotions.” **Nature reviews. Neuroscience**, v. 15, n. 3, p. 170–80, 2014.

KUMAR, A.; WELTI, D.; ERNST, R. R. “NMR Fourier zeugmatography”. **Journal of Magnetic Resonance (1969)**, v. 18, n. 1, p. 69–83, 1975.

LANG, P. J. “A bio-informational theory of emotional imagery.” **Psychophysiology**, v. 16, n. 6, p. 495–512, 1979.

LANGERS, D. R. M.; VAN DIJK, P. “Mapping the Tonotopic Organization in Human Auditory Cortex with Minimally Salient Acoustic Stimulation”. **Cerebral Cortex**, v. 22, n. 9, p. 2024–2038, 2012.

LARTILLOT, O.; TOIVIAINEN, P. “Mir in Matlab ( Ii ): a Toolbox for Musical Feature Extraction From Audio”. **Flux**, n. Ii, p. 237–244, 2007.

LARTILLOT, O.; TOIVIAINEN, P.; EEROLA, T. “A Matlab Toolbox for Music Information Retrieval”. **Data Analysis Machine Learning and Applications, Studies in Classification, Data Analysis, and Knowledge Organization**, p. 261–268,

2008.

LAURIER, C. et al. “Exploring Relationships between Audio Features and Emotion in Music”. **Triennial Conference of European Society for the Cognitive Sciences of Music**, n. Escom, p. 260–264, 2009.

LAUTERBUR, P. C. “Image Formation by Induced Local Interactions: Examples Employing Nuclear Magnetic Resonance”. **Nature**, v. 242, n. 5394, p. 190–191, 16 mar. 1973.

LEAVER, A. M.; RAUSCHECKER, J. P. “Cortical Representation of Natural Complex Sounds: Effects of Acoustic Features and Auditory Object Category”. **Journal of Neuroscience**, v. 30, n. 22, p. 7604–7612, 2010.

LEHNE, M.; ROHRMEIER, M.; KOELSCH, S. “Functional neuroimaging of tension: A study with music”. **Social Cognitive and Affective Neuroscience**, 2013.

LEWIS, M.; HAVILAND-JONES, J. M.; BARRETT, L. F. **Handbook of emotions**. 3 edition ed. New York: The Guildford Press, 2008.

LIU, T. T.; NALCI, A.; FALAHPOUR, M. “The global signal in fMRI: Nuisance or Information?” **NeuroImage**, v. 150, p. 213–229, 15 abr. 2017.

MASON, M. F. et al. “Wandering minds: the default network and stimulus-independent thought.” **Science (New York, N.Y.)**, v. 315, n. 5810, p. 393–5, 19 jan. 2007.

MITTERSCHIFFTHALER, M. T. et al. “A functional MRI study of happy and sad affective states induced by classical music.” **Human brain mapping**, v. 28, n. 11, p. 1150–62, nov. 2007.

MOEREL, M.; DE MARTINO, F.; FORMISANO, E. “Processing of natural sounds in human auditory cortex: tonotopy, spectral tuning, and relation to voice sensitivity.” **The Journal of neuroscience: the official journal of the Society for Neuroscience**, v. 32, n. 41, p. 14205–16, 10 out. 2012.

MOEREL, M.; DE MARTINO, F.; FORMISANO, E. “An anatomical and functional topography of human auditory cortical areas”. **Frontiers in Neuroscience**, v. 8, n. 8 JUL, p. 1–14, 2014.

MOEREL, M. M. L. **Encoding of natural sounds in the human brain**. [s.l.] Maastricht University, 2013.

MURPHY, K. et al. “The impact of global signal regression on resting state correlations: are anti-correlated networks introduced?” **NeuroImage**, v. 44, n. 3, p. 893–905, 1 fev. 2009.

NASELARIS, T. et al. “Bayesian Reconstruction of Natural Images from Human Brain Activity”. **Neuron**, v. 63, n. 6, p. 902–915, 2009.

NASELARIS, T. et al. “Encoding and decoding in fMRI”. **NeuroImage**, v. 56, n. 2, p. 400–410, 2011.

NISHIMOTO, S. et al. “Reconstructing visual experiences from brain activity evoked by natural movies”. **Current Biology**, v. 21, n. 19, p. 1641–1646, 11 out. 2011.

NORMAN-HAIGNERE, S.; KANWISHER, N. G.; MCDERMOTT, J. H. “Distinct Cortical Pathways for Music and Speech Revealed by Hypothesis-Free Voxel Decomposition”. **Neuron**, v. 88, n. 6, p. 1281–1296, dez. 2015.

PARK, M. et al. “Personality traits modulate neural responses to emotions

expressed in music”. **Brain Research**, v. 1523, p. 68–76, 2013.

PATTERSON; GRIFFITHS. “The Processing of Temporal Pitch and Melody Information in Auditory Cortex”. **Neuron**, v. 36, n. 4, p. 767–776, 2002.

PAULI, W. “Zur Frage der theoretischen Deutung der Satelliten einiger Spektrallinien und ihrer Beeinflussung durch magnetische Felder”. **Die Naturwissenschaften**, v. 12, n. 37, p. 741–743, set. 1924.

PEDREGOSA, F. AND VAROQUAUX, G. AND GRAMFORT, A. AND MICHEL, V. et al. “Scikit-learn: Machine Learning in Python”. **Journal of Machine Learning Research**, n. 12, p. 2825–2830, 2011.

PRICE, C.; THIERRY, G.; GRIFFITHS, T. “Speech-specific auditory processing: where is it?” **Trends in Cognitive Sciences**, v. 9, n. 6, p. 271–276, 2005.

RABI, I. I. et al. “The Molecular Beam Resonance Method for Measuring Nuclear Magnetic Moments. The Magnetic Moments of  $\text{Li } 6 \text{ } ^3$ ,  $\text{Li } 7 \text{ } ^3$  and  $\text{F } 19 \text{ } ^9$ ”. **Physical Review**, v. 55, n. 6, p. 526–535, 15 mar. 1939.

SAAD, Z. S. et al. “Trouble at rest: how correlation patterns and group differences become distorted after global signal regression.” **Brain connectivity**, v. 2, n. 1, p. 25–32, 10 jan. 2012.

SAENZ, M.; LANGERS, D. R. M. “Tonotopic mapping of human auditory cortex”. **Hearing Research**, v. 307, p. 42–52, 2014.

SALIMPOOR, V. N. et al. “Interactions Between the Nucleus Accumbens and Auditory Cortices Predict Music Reward Value”. **Science**, v. 340, n. 6129, p. 216–219, 11 abr. 2013.

SALIMPOOR, V. N.; ZATORRE, R. J. “Neural interactions that give rise to musical pleasure.” **Psychology of Aesthetics, Creativity, and the Arts**, v. 7, n. 1, p. 62–75, 2013.

SANTORO, R. et al. “Encoding of Natural Sounds at Multiple Spectral and Temporal Resolutions in the Human Auditory Cortex”. **PLoS Computational Biology**, v. 10, n. 1, p. e1003412, 2 jan. 2014.

SANTORO, R. et al. “Reconstructing the spectro-temporal modulations of real-life sounds from fMRI response patterns”. **Pnas**, v. 114, n. 18, 2017.

SCHACHTER, S.; SINGER, J. “Cognitive, social, and physiological determinants of emotional state.” **Psychological review**, v. 69, n. 5, p. 379, 1962.

SCHERER, K. R. “Toward a dynamic theory of emotion: The component process model of affective states”. **Geneva Studies in Emotion and Communication**, v. 1, n. 1, p. 1–98, 1987.

SCHERER, K. R. “Which Emotions Can be Induced by Music ? What Are the Underlying Mechanisms ? And How Can We Measure Them ?” v. 33, n. 3, p. 239–251, 2004.

SCHERER, K. R. “What are emotions? And how can they be measured?” **Social Science Information**, v. 44, n. 4, p. 695–729, 2005.

SCHÖLVINCK, M. L. et al. “Neural basis of global resting-state fMRI activity.” **Proceedings of the National Academy of Sciences of the United States of America**, v. 107, n. 22, p. 10238–43, 1 jun. 2010.

SCHÖNWIESNER, M. et al. “Parcellation of Human and Monkey Core

Auditory Cortex with fMRI Pattern Classification and Objective Detection of Tonotopic Gradient Reversals.” **Cerebral cortex (New York, N.Y. : 1991)**, p. 1–12, 2014.

SCHÖNWIESNER, M.; VON CRAMON, D. Y.; RÜBSAMEN, R. “Is It Tonotopy after All?” **NeuroImage**, v. 17, n. 3, p. 1144–1161, 2002.

SHANNON, C. E. “A Mathematical Theory of Communication”. **The Bell System Technical Journal**, v. 27, n. 1, p. 379–423, 623–656, 1948.

SMITH, S. M. et al. “Advances in functional and structural MR image analysis and implementation as FSL.” **NeuroImage**, v. 23 Suppl 1, p. S208-19, jan. 2004.

STANSBURY, D.; NASELARIS, T.; GALLANT, J. “Natural Scene Statistics Account for the Representation of Scene Categories in Human Visual Cortex”. **Neuron**, v. 79, p. 1025–1034, 2013.

STRIEM-AMIT, E.; HERTZ, U.; AMEDI, A. “Extensive Cochleotopic Mapping of Human Auditory Cortical Fields Obtained with Phase-Encoding fMRI”. **PLoS ONE**, v. 6, n. 3, p. e17832, 2011.

THOMAS, J. M. et al. “Population receptive field estimates of human auditory cortex”. **NeuroImage**, v. 105, p. 428–439, 2015.

TOIVIAINEN, P. et al. “Capturing the musical brain with Lasso: Dynamic decoding of musical features from fMRI data”. **NeuroImage**, v. 88, p. 170–180, 19 dez. 2014.

TROST, W. et al. “Mapping Aesthetic Musical Emotions in the Brain.” **Cerebral cortex (New York, N.Y. : 1991)**, p. 1–15, 15 dez. 2011.

TZANETAKIS, G.; COOK, P. “Musical genre classification of audio signals”. **IEEE Transactions on Speech and Audio Processing**, v. 10, n. 5, p. 293–302, 2002.

UHLENBECK, G. E.; GOUDSMIT, S. “Ersetzung der Hypothese vom unmechanischen Zwang durch eine Forderung bezüglich des inneren Verhaltens jedes einzelnen Elektrons”. **Die Naturwissenschaften**, v. 13, n. 47, p. 953–954, nov. 1925.

VAROQUAUX, G. et al. “Assessing and tuning brain decoders: Cross-validation, caveats, and guidelines”. **NeuroImage**, v. 145, p. 166–179, jan. 2017.

VAROQUAUX, G.; THIRION, B. “How machine learning is shaping cognitive neuroimaging”. **Giga Science**, n. 3, p. 28, 2014.

VIINIKAINEN, M.; KÄTSYRI, J.; SAMS, M. “Representation of perceived sound valence in the human brain”. **Human Brain Mapping**, v. 33, p. 2295–2305, 2012.

WARREN, J. D. et al. “Analyzing Pitch Chroma and Pitch Height in the Human Brain”. **Proceedings of the National Academy of Sciences of the United States of America**, v. 100, n. 17, p. 10038–10042, 2003.

ZATORRE, R. J.; BELIN, P.; PENHUNE, V. B. “Structure and function of auditory cortex music and speech”. **Trends in Cognitive Sciences**, v. 6, n. 1, p. 37–46, 2002.

ZENTNER, M.; GRANDJEAN, D.; SCHERER, K. R. “Emotions evoked by the sound of music: characterization, classification, and measurement.” **Emotion (Washington, D.C.)**, v. 8, n. 4, p. 494–521, ago. 2008.

## 8 Apêndice

### 8.1 Correção do sinal global

A correção do sinal global como um passo de pré-processamento está até hoje um tópico controverso (GLASSER et al., 2016; LIU; NALCI; FALAHPOUR, 2017; MURPHY et al., 2009). O sinal global é definido pela média de todos os voxels e por isso reflete variações generalizadas, que possam ser introduzidas, por exemplo, por efeitos fisiológicos. Nesse estudo, a correção do sinal global foi aplicada com o intuito de normalizar o sinal BOLD, já que o sinal global mostrou uma diferença no fator experimental da tarefa, descrito a seguir. Para a tarefa de sentir a emoção, o sinal global teve uma variação maior e, ao mesmo tempo, teve uma correlação maior entre todos os *voxels* do cérebro. A Figura 8-1 mostra essa diferença para um sujeito típico. As correlações foram calculadas a partir do dado pré-processado (suavizado espacialmente com FWHM = 8 mm), porém sem ter aplicado a correção do sinal global. A Figura 8-2 mostra como esse efeito pode ser corrigido quando se aplica a correção do sinal global. A Figura 8-3 mostra o resultado do teste *t* pareado sobre os primeiros quatro sujeitos e 8 *medleys* antes e depois da correção.

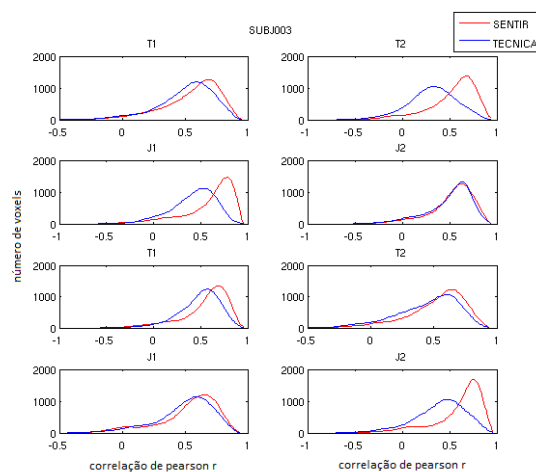


Figura 8-1: Distribuição das correlações do sinal global com todos os voxels por medley, antes da correção do sinal global. Para este participante, as correlações de 34.780 voxels são visualizadas. A tarefa de sentir a emoção mostra correlações mais altas com o sinal global (em vermelho) do que a tarefa técnica (em azul).

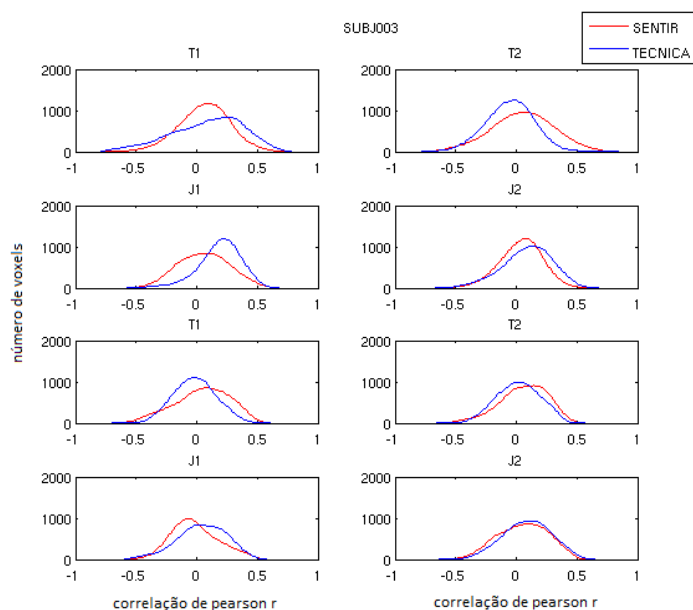


Figura 8-2: Distribuição das correlações depois da correção do sinal global (sinal global com todos os voxels por medley). A correção centraliza a média das correlações e remove o efeito sistemático da tarefa.

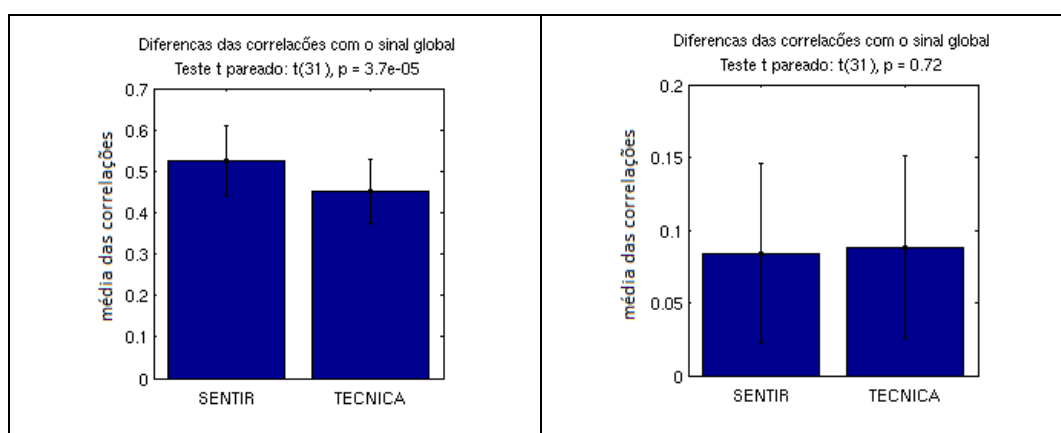


Figura 8-3: Diferenças entre a tarefa de sentir e não sentir (tarefa técnica) no sinal global. Antes da correção do sinal global (lado esquerdo) a tarefa sentir mostra correlações maiores (teste t pareado sobre a média das correlações por sujeito e medley:  $t(31)$ ,  $p=3,7 \cdot 10^{-5}$ ). Após correção (lado direito), não existem mais diferenças estatísticas. Além disso, as médias das correlações estão mais próximas ao zero.

O significado do sinal global ainda é discutido amplamente na literatura (CHAI et al., 2012; CHEN et al., 2012; FOX et al., 2009; MURPHY et al., 2009; SAAD et al., 2012; SCHÖLVINCK et al., 2010). No presente contexto, foi observado o fato de que o sinal global mostra uma diferença entre as tarefas. Além disso, a remoção do sinal

global se mostrou benéfico para a análise principal, a codificação das características musicais no córtex auditivo, como mostra a Figura 8-4.

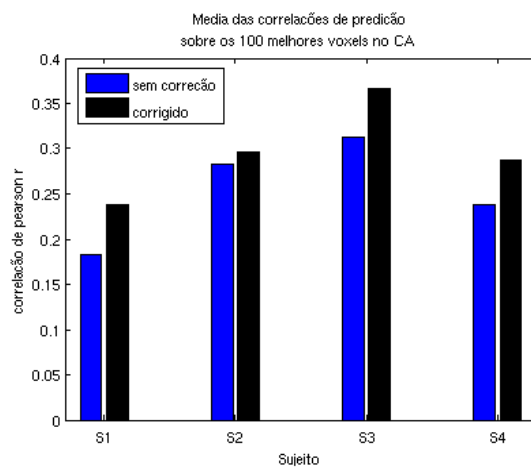


Figura 8-4: Efeito da remoção do sinal global na codificação. Com a correção, as correlações de predição aumentaram em todos os sujeitos.

## 8.2 Telas da pontuação

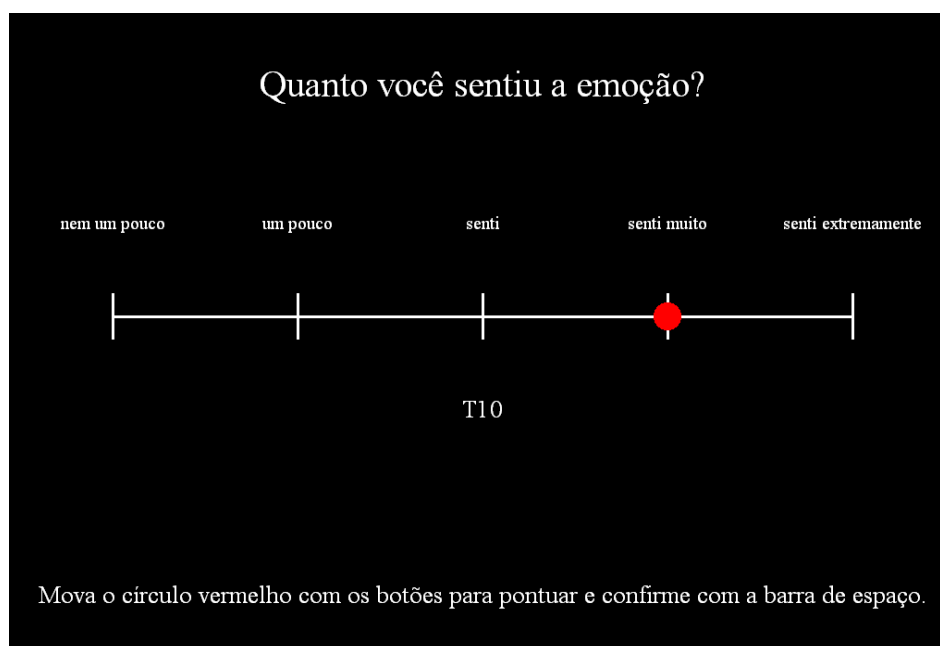


Figura 8-5: Tela de pontuação: Intensidade da emoção



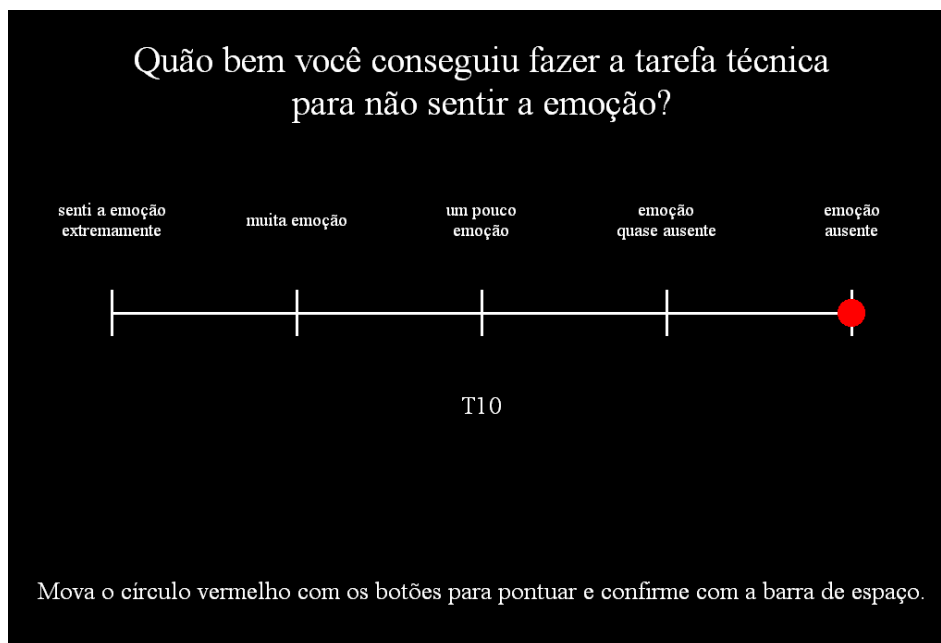


Figura 8-6: Tela de pontuação: Tarefa técnica para não sentir a emoção

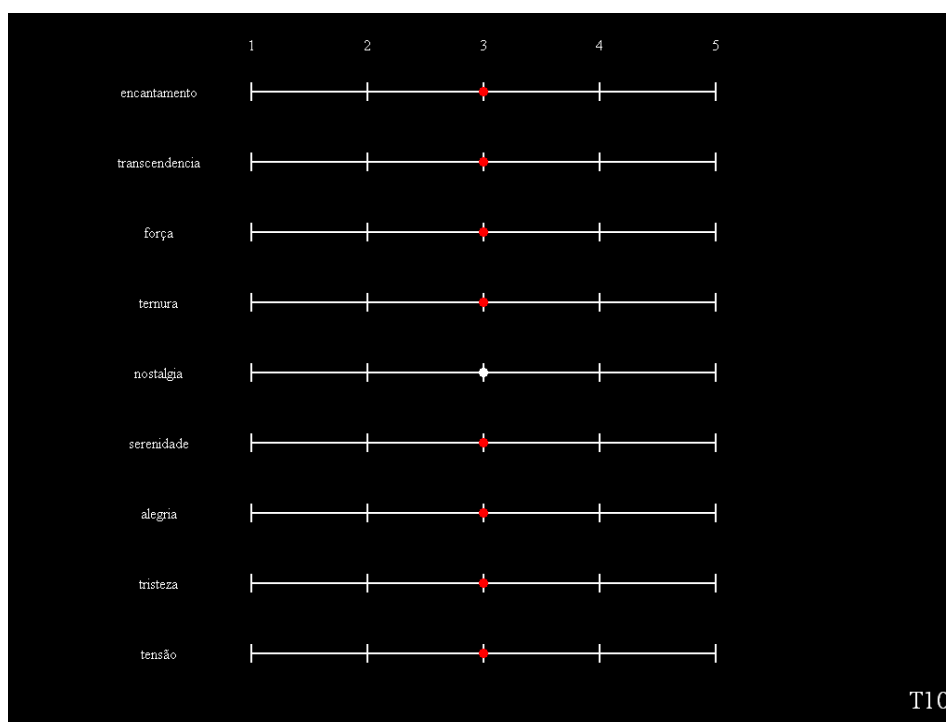


Figura 8-7: Escala de GEMS de 9 itens antes de pontuar. A pontuação foi realizada com o mouse marcando na posição desejada de cada escala. O tempo para pontuar era livre. Para confirmar e pontuar a próxima música. A barra de espaço confirmava a pontuação.

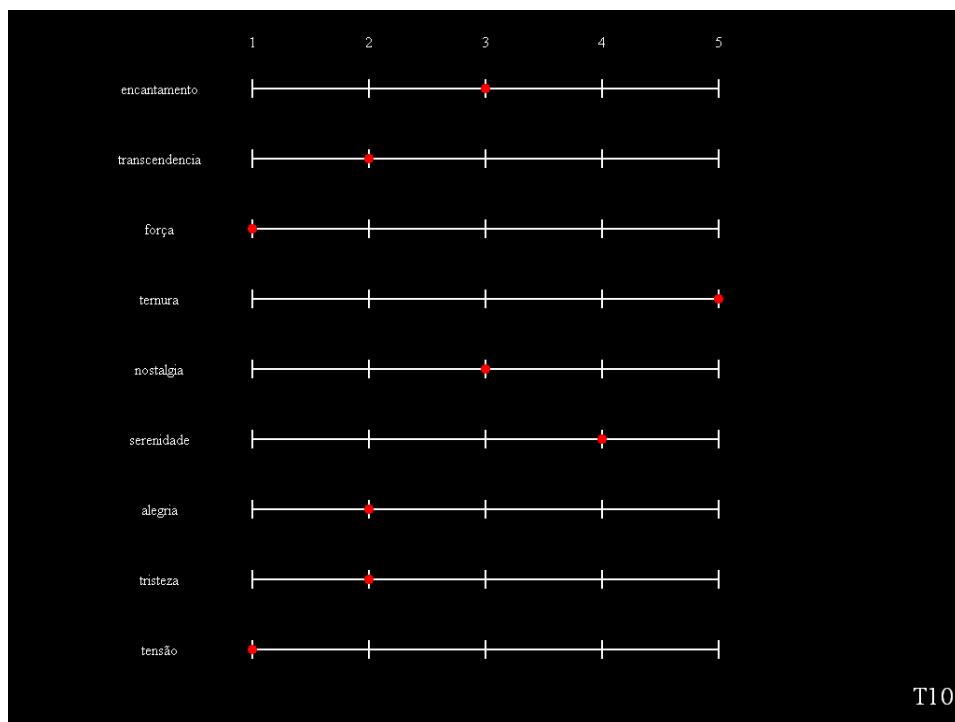


Figura 8-8: Escala de GEMS com a pontuação realizada para uma música exemplar.

### 8.3 Questionário pós-RM

SUBJ: 000

Data: dd/mm/aaaa

#### Entrevista pós – experimento – RM

Por favor, responda as perguntas abaixo sobre o teste na ressonância magnética “**EmoCode**” e marque sua resposta com uma cruz usando a escala abaixo da pergunta:

Quão difícil foi a tarefa de sentir a emoção alegria (joyful activation) para você?				
muito fácil	fácil	nem fácil, nem difícil	difícil	muito difícil
Quão difícil foi a tarefa de sentir a emoção ternura para você?				
muito fácil	fácil	nem fácil, nem difícil	difícil	muito difícil
Quão difícil foi a tarefa técnica com as músicas de alegria (joyful activation) para você?				
muito fácil	fácil	nem fácil, nem difícil	difícil	muito difícil
Quão difícil foi a tarefa técnica com as músicas de ternura para você?				
muito fácil	fácil	nem fácil, nem difícil	difícil	muito difícil
Quão concentrado você estava para a tarefa de sentir a emoção alegria (joyful activation)?				
nem um pouco	pouco	concentrado	bastante	extremamente
Quão concentrado você estava para a tarefa de sentir a emoção ternura?				
nem um pouco	pouco	concentrado	bastante	extremamente
Quão concentrado você estava para a tarefa técnica com as músicas de alegria?				
nem um pouco	pouco	concentrado	bastante	extremamente
Quão concentrado você estava para a tarefa técnica com as músicas de ternura?				
nem um pouco	pouco	concentrado	bastante	extremamente
Quão motivado você estava para fazer as tarefas?				
nem um pouco	pouco	motivado	bastante	extremamente
Quão entediado (de saco cheio) você estava durante o experimento?				
nem um pouco	pouco	entediado	bastante	extremamente

Qual estratégia você usou para “sentir a emoção”de ternura?

Qual estratégia você usou para “sentir a emoção”de alegria (joyful activation)?

Qual estratégia você usou para fazer a tarefa técnica com as músicas de ternura?

Qual estratégia você usou para fazer a tarefa técnica com as músicas de alegria (joyful activation)?

SUBJ: 000

Data: dd/mm/aaaa

---

**Alguma coisa te incomodou? (ex. ouvir os trechos varias vezes, responder às perguntas, barulho da maquina, fone de ouvido)**

Sim Não

Se sim, o quê?

**Você se mexeu enquanto escutava as músicas (p.ex., acompanhando os ritmos com o pé ou a mão)?**

Sim Não Às vezes Por favor, explique:

**Você conseguiu deixar os olhos abertos enquanto escutava as músicas?**

Sim Não Às vezes Por favor, explique:

**O volume estava bom, você ouviu os trechos bem?**

Sim Não Se não, por favor, explique:

**O que foi mais difícil para você no experimento? Por favor, explique:**

*Figura 8-9: Questionário preenchido depois de cada sessão de RM.*